

Impact of Gene Expression Profiling Tests on Breast Cancer Outcomes

Prepared for:

Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-02-0018

Prepared by:

The Johns Hopkins University Evidence-based Practice Center, Baltimore, MD

Investigators

Luigi Marchionni, M.D., Ph.D.
Renee F. Wilson, M.Sc.
Spyridon S. Marinopoulos, M.D., M.B.A.
Antonio C. Wolff, M.D.
Giovanni Parmigiani, M.D.
Eric B. Bass, M.D., M.P.H.
Steven N. Goodman, M.D., M.H.S., Ph.D.

This report is based on research conducted by the Johns Hopkins University Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-02-0018). The findings and conclusions in this document are those of the author(s), who are responsible for its content, and do not necessarily represent the views of AHRQ. No statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help clinicians, employers, policymakers, and others make informed decisions about the provision of health care services. This report is intended as a reference and not as a substitute for clinical judgment.

This report may be used, in whole or in part, as the basis for the development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

This document is in the public domain and may be used and reprinted without permission except those copyrighted materials noted for which further reproduction is prohibited without the specific permission of copyright holders.

Suggested Citation:

Marchionni L, Wilson RF, Marinopoulos SS, Wolff AC, Parmigiani G, Bass EB, Goodman SN. Impact of Gene Expression Profiling Tests on Breast Cancer Outcomes. Evidence Report/Technology Assessment No. 160. (Prepared by The Johns Hopkins University Evidence-based Practice Center under contract No. 290-02-0018). AHRQ Publication No. 08-E002. Rockville, MD: Agency for Healthcare Research and Quality. January 2008.

The investigators have no relevant financial interests in the report. The investigators have no employment, consultancies, honoraria, or stock ownership or options, or royalties from any organization or entity with a financial interest or financial conflict with the subject matter discussed in the report.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-Based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The Centers for Disease Control and Prevention (CDC) requested and provided funding for this report. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To bring the broadest range of experts into the development of evidence reports and health technology assessments, AHRQ encourages the EPCs to form partnerships and enter into collaborations with other medical and research organizations. The EPCs work with these partner organizations to ensure that the evidence reports and technology assessments they produce will become building blocks for health care quality improvement projects throughout the Nation. The reports undergo peer review prior to their release.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality.

We welcome comments on this evidence report. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by e-mail to epc@ahrq.gov.

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Julie Louise Gerberding, M.D., M.P.H.
Director
Centers for Disease Control and Prevention

Gurvaneet Randhawa, M.D., M.P.H.
EPC Program Task Order Officer
Agency for Healthcare Research and Quality

Beth Collins Sharp, Ph.D., R.N.
Director, EPC Program
Agency for Healthcare Research and Quality

Acknowledgments

The Evidence-based Practice Center thanks Michael Oladubu, D.D.S. and Allison Jonas, for their assistance with literature searching and database management, and project organization; Aly Shogan for her assistance in completing the sections on economics; Brenda Zacharko for her assistance with budget matters, and for her assistance with final preparations of the report. The Center also wishes to thank Gurveet Randhawa, M.D., M.P.H., AHRQ Task Order Officer, for his efforts in guiding this project and coordination with the CDC EGAPP group.

Structured Abstract

Objective: To assess the evidence that three marketed gene expression-based assays improve prognostic accuracy, treatment choice, and health outcomes in women diagnosed with early stage breast cancer.

Data Sources: MEDLINE[®], EMBASE, the Cochrane databases, test manufacturer Web sites, and information provided by manufacturers.

Review Methods: We evaluated the evidence for three gene expression assays on the market; *Oncotype DX*[™], MammaPrint[®] and the Breast Cancer Profiling (BCP or H/I ratio) test, and for gene expression signatures underlying the assays. We sought evidence on: (a) analytic performance of tests; (b) clinical validity (i.e., prognostic accuracy and discrimination); (c) clinical utility (i.e., prediction of treatment benefit); (d) harms; and (e) impact on clinical decision making and health care costs.

Results: Few papers were found on the analytic validity of the *Oncotype DX* and MammaPrint tests, but these showed reasonable within-laboratory replicability. Pre-analytic issues related to sample storage and preparation may play a larger role than within-laboratory variation. For clinical validity, studies differed according to whether they examined the actual test that is currently being offered to patients or the underlying gene signature. Almost all of the *Oncotype DX* evidence was for the marketed test, the strongest validation study being from one arm of a randomized controlled trial (NSABP-14) with a clinically homogeneous population. This study showed that the test, added in a clinically meaningful manner to standard prognostic indices. The MammaPrint signature and test itself was examined in studies with clinically heterogeneous populations (e.g., mix of ER positivity and tamoxifen treatment) and showed a clinically relevant separation of patients into risk categories, but it was not clear exactly how many predictions would be shifted across decision thresholds if this were used in combination with traditional indices. The BCP test itself was examined in one study, and the signature was tested in a variety of formulations in several studies. One randomized controlled trial provided high quality retrospective evidence of the clinical utility of *Oncotype DX* to predict chemotherapy treatment benefit, but evidence for clinical utility was not found for MammaPrint or the H/I ratio. Three decision analyses examined the cost-effectiveness of breast cancer gene expression assays, and overall were inconclusive.

Conclusions: *Oncotype DX* is furthest along the validation pathway, with strong retrospective evidence that it predicts distant spread and chemotherapy benefit to a clinically relevant extent over standard predictors, in a well-defined clinical subgroup with clear treatment implications. The evidence for clinical implications of using MammaPrint was not as clear as with *Oncotype DX*, and the ability to predict chemotherapy benefit does not yet exist. The H/I ratio test requires further validation. For all tests, the relationship of predicted to observed risk in different populations still needs further study, as does their incremental contribution, optimal implementation, and relevance to patients on current therapies.

Contents

| | |
|--|----------|
| Executive Summary | 1 |
| Evidence Report | 9 |
| Chapter 1. Introduction | 11 |
| Breast Cancer | 11 |
| Gene expression profiling..... | 12 |
| Breast Cancer Assays on the Market | 13 |
| RT-PCR..... | 14 |
| Microarrays..... | 15 |
| Sources of Variability in Gene Expression Analysis..... | 16 |
| Objectives of the Evidence Report | 17 |
| Structured Approach to Assessment of the Questions..... | 18 |
| Chapter 2. Methods..... | 21 |
| Recruitment of Technical Experts and Peer Reviewers..... | 21 |
| Key Questions..... | 21 |
| Literature Search Methods..... | 21 |
| Sources..... | 22 |
| Search terms and strategies..... | 22 |
| Organization and tracking of literature search..... | 23 |
| Title Review..... | 23 |
| Abstract Review..... | 23 |
| Inclusion and exclusion criteria | 23 |
| Article Inclusion/Exclusion | 24 |
| Data Abstraction | 26 |
| Quality Assessment..... | 26 |
| Data Synthesis..... | 27 |
| Data Entry and Quality Control | 27 |
| Grading of the Evidence | 27 |
| Peer Review | 27 |
| Chapter 3. Results | 29 |
| Key Question 1. What is the direct evidence that gene expression profiling tests in women diagnosed with breast cancer, or any specific subset of this population, lead to improvement in outcomes?..... | 29 |
| Key Question 2. What are the sources of and contributions to analytic validity in these gene expression-based prognostic estimators for women diagnosed with breast cancer?..... | 29 |
| <i>Oncotype DX</i> TM | 30 |
| MammaPrint® | 34 |
| H/I Ratio..... | 36 |
| Key Question 3. What is the clinical validity of gene expression profiling tests in women diagnosed with breast cancer? | 38 |

| | |
|---|-----|
| <i>Oncotype DX</i> | 38 |
| MammaPrint | 39 |
| H/I Ratio..... | 41 |
| Key Question 4. What is the clinical utility of these tests? | 45 |
| <i>Oncotype DX</i> | 46 |
| MammaPrint | 52 |
| H/I Ratio..... | 54 |
| Ongoing Studies..... | 55 |
| TAILORx..... | 55 |
| MINDACT..... | 55 |
| Other Relevant Studies | 55 |
| Studies Excluded Upon Complete Review | 57 |
| | |
| Chapter 4. Discussion | 87 |
| <i>Oncotype DX</i> | 88 |
| Analytic validity..... | 88 |
| Clinical validity..... | 89 |
| Clinical utility | 90 |
| Questions regarding the clinical validity and utility of the <i>Oncotype DX</i> assay | 93 |
| MammaPrint | 93 |
| Analytic validity..... | 94 |
| Clinical validity..... | 94 |
| Clinical utility | 95 |
| H/I Ratio Signature and Breast Cancer Profiling (BCP) | 96 |
| General Comments on Analytic Validity and Laboratory Quality Control..... | 96 |
| Overall implications and recommendations..... | 97 |
| Assay validation..... | 97 |
| Potential for scale problems..... | 97 |
| Genetic variability and gene expression | 98 |
| The need for databases, reproducibility, and standards | 98 |
| Where is the field going? | 98 |
| “Comparative effectiveness” studies | 99 |
| Conclusion | 99 |
| | |
| References and Included Studies | 101 |

Tables

| | |
|--|----|
| Table 1. Description of the three gene expression profile assays | 59 |
| Table 2. Successful assays, <i>Oncotype DX</i> | 62 |
| Table 3. Variability and reproducibility, <i>Oncotype DX</i> | 63 |
| Table 4. Analytic validity, <i>Oncotype DX</i> | 64 |
| Table 5. RT-PCR vs. IHC comparison assays, <i>Oncotype DX</i> | 65 |
| Table 6. Successful assays, MammaPrint..... | 67 |
| Table 7. Reproducibility, MammaPrint | 68 |
| Table 8. Analytic validity, MammaPrint | 69 |

| | |
|---|----|
| Table 9. Successful assays, two-gene signature and H/I ratio assays | 70 |
| Table 10. Reproducibility, two-gene signature and H/I ratio assay | 71 |
| Table 11. RT-PCR vs. IHC comparison assays, two-gene signature and H/I ratio assay | 72 |
| Table 12. Clinical validity, <i>Oncotype DX</i> | 73 |
| Table 13. Risk classification of <i>Oncotype DX</i> against the St. Gallen criteria | 75 |
| Table 14. Risk classification of <i>Oncotype DX</i> against the 2004 NCCN guidelines..... | 75 |
| Table 15. Risk classification of <i>Oncotype DX</i> against the Adjuvant! Guidelines..... | 75 |
| Table 16. Clinical Validity, MammaPrint and 70-gene signature | 76 |
| Table 17. MammaPrint compared with traditional composite risk markers..... | 79 |
| Table 18. Clinical Validity, two-gene signature and H/I ratio assays | 80 |
| Table 19. Clinical Utility, <i>Oncotype DX</i> | 83 |
| Table 20. Comparison of economic studies..... | 85 |
| Table 21. Clinical Utility, two-gene signature and H/I ratio | 86 |

Figures

| | |
|---|----|
| Figure 1. Increasing complexity of information from genome to transcriptome and proteome: gene expression analysis focuses on the analysis of the transcriptome..... | 12 |
| Figure 2. Quantitative RT-PCR..... | 15 |
| Figure 3. Schematic model for microarray hybridizations..... | 16 |
| Figure 4. Summary of literature search and review process (number of articles)..... | 25 |

Appendixes

| |
|--|
| Appendix A: List of Acronyms |
| Appendix B: Glossary |
| Appendix C: Description of Genes |
| Appendix D: Technologies |
| Appendix E: Technical Experts and Peer Reviewers |
| Appendix F: Detailed Electronic Database Search Strategies |
| Appendix G: Review Forms |
| Appendix H: Excluded Articles |
| Appendix I: Evidence Tables |

Appendixes and Evidence Tables for this report are provided electronically at
<http://www.ahrq.gov/downloads/pub/evidence/pdf/brcancergene/brcangene.pdf>.

Executive Summary

Introduction

Breast cancer is the most commonly diagnosed cancer in women. This tumor is the second leading cause of cancer-related deaths in women in the United States, with approximately 178,000 new cases and 40,000 deaths expected among U.S. women in 2007. Treatment for breast cancer usually involves surgery to remove the tumor and involved lymph nodes. Frequently, surgery is followed by radiation therapy (in case of breast conservation or in women with large tumors or many involved lymph nodes), endocrine therapy (for essentially all women with tumors that express the estrogen receptor (ER-positive)), and/or chemotherapy (for women having a high risk for a poor outcome such as those with large tumors, involved lymph nodes, advanced disease, or inflammatory breast cancer). More than three-quarters of patients are expected to survive with this multi-modality approach.

Gene expression profiling has been proposed as an approach to address this issue in clinical settings, and three breast cancer gene expression assays are now available in the U.S. The *Oncotype DX*TM Breast Cancer Assay, the MammaPrint[®] Test, and the Breast Cancer Profiling test (BCP or H/I ratio). MammaPrint is based on the use of microarray technology, while the other two assays are based on the reverse transcriptase polymerase chain reaction (RT-PCR). All of these tests combine the measurements of gene expression levels within the tumor to produce a number associated with the risk of distant disease recurrence. These tests aim to improve on risk stratification schemes based on clinical and pathologic factors currently used in clinical practice. As therapeutic decisions are based on risk estimates, tests that improve such estimates have the potential to affect clinical outcome in breast cancer patients by either avoiding unnecessary chemotherapy and its attendant morbidity or by employing it where it might not otherwise have been used, thereby reducing recurrence risk.

The literature was searched for evidence about the use of gene expression profiling in breast cancer. Our analytical framework for reporting the results distinguishes between the assays, as they are offered to patients, and the underlying signatures, which comprise the genes whose expression is measured. This measurement of expression can be done in a number of ways that may not be identical to the procedures used for the marketed test, producing an unknown number of different predictions. We also distinguish between developmental and validation studies.

Methods

Working with the Agency for Healthcare Research and Quality (AHRQ), the Centers for Disease Prevention and Control (CDC), the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) working group, and members of a technical expert panel, we formulated four key questions, and addressed them on the basis of the evidence available about the specific assays and the underlying gene expression signatures. The original set of key questions was refined to focus primarily on two gene expression profiling tests: *Oncotype DX* (Genomic Health, Inc.) and MammaPrint (Agendia). During the course of the evaluation, a third gene expression profiling test came to our attention, the H/I ratio test based on the two-gene signature (AviaraDX/Quest Diagnostics, Inc.), and was thus investigated. We searched and retrieved

studies in MEDLINE[®], EMBASE, and the Cochrane databases (1990-2006). We supplemented this search with recent publications that appeared after the time period initially considered in the systematic search, and about the two-gene test (H/I ratio). We also searched for relevant documents on the Food and Drug Administration's web site, and solicited additional documentation from the companies offering the tests. The systematic searches yielded a total of 12983 citations. Specific inclusion and exclusion criteria were developed and pairs of readers reviewed each title; the same procedure was used to review selected abstracts. We identified 63 studies for full text review. We developed tables to summarize each article. Initial data were abstracted by investigators and entered directly into evidence tables. Quality and consistency of the abstracted data was then evaluated by a second reviewer, and a senior investigator examined all reviews to identify potential problems with data abstraction. These were discussed at meetings of group members. A system of random data checks was applied to ensure data abstraction accuracy.

Results

Literature on Key Questions

Key Question 1. What is the direct evidence that gene expression profiling tests in women diagnosed with breast cancer (or any specific subset of this population) lead to improvement in outcomes?

Direct evidence was defined as a study where the primary intervention is the use of a prognostic test (with therapeutic decisionmaking directed by the result) and the outcomes are patient morbidity, mortality and/or quality of life. No direct evidence was found in the published data on improvement of patients' outcomes due to such testing in women diagnosed with breast cancer, nor were there any randomized studies using the tests' predictions to manage patients. However, as described under Key Questions 3 and 4, some of the tests' supporting evidence was derived from past randomized controlled trials (RCTs) with prospectively gathered patient samples, giving them strong evidential value. Two ongoing RCTs, TAILORx and MINDACT (using *Oncotype DX*, and *MammaPrint* respectively), will provide further evidence allowing almost direct inference about the impact on patient outcomes.

Key Question 2. What are the sources of and contributions to analytic validity in these two gene expression-based prognostic estimators for women diagnosed with breast cancer?

In the field of gene expression there are no "gold standards" outside the technologies used in the tests under study, i.e., microarrays and RT-PCR. Consequently, a definitive evaluation of the analytic validity of expression-based tests is difficult. Evidence about operational characteristics was partial and limited to a few publications. A 2007 paper by Cronin and colleagues, on the analytic validity of *Oncotype DX* was the most detailed study for any of these tests so far, showing good performance for a number of analytic components of the assay. Data about the sources and contributions to variability of the tests and about their reproducibility was generally limited to analyses of few samples, and thus a complete evaluation of the impact of such variability on risk assessment was not available. Partial evidence about analytic validity was provided in the percentage of subjects whose samples were successfully analyzed with these tests, and those numbers were fairly good. Continuous monitoring of laboratory procedures and

Careful evaluation of the quality of the submitted specimens are major factors affecting test reliability.

Key Question 3. What is the clinical validity of these tests in women diagnosed with breast cancer?

- a. How well does this testing predict recurrence rates for breast cancer compared to standard prognostic approaches? Specifically, how much do these tests add to currently known factors or combination indices that predict the probability of breast cancer recurrence, (e.g., tumor type or stage, age, ER, and human epidermal growth factor receptor 2 (HER-2) status)?
- b. Are there any other factors, which may not be components of standard predictors of recurrence (e.g., race/ethnicity or adjuvant therapy), that affect the clinical validity of these tests, and thereby generalizability of results to different populations?

Clinical validity is defined as the degree to which a test accurately predicts the risk of an outcome (i.e., calibration), as well as its ability to separate patients with different outcomes into separate risk classes (discrimination). Clinical validity was documented to some degree for all three gene expression signatures. *Oncotype DX* was validated on a homogenous population of lymph node negative, ER positive patients all treated with tamoxifen, derived from an arm of an RCT, the National Surgical Adjuvant Breast and Bowel Project (NSABP-14). *MammaPrint*, on the other hand, was validated on samples from a clinical series with a wide range of clinical and treatment characteristics, and sometimes it was the signature and not the *MammaPrint* test itself that was validated. Data that made clear the incremental value of the test over standardized risk predictors using classical clinical factors, in the form of risk reclassification tables, was limited to *Oncotype DX* in one population, and for one of those predictors (*Adjuvant! Online* for *MammaPrint*). The evidence behind the two-gene test is quite heterogeneous, in that the specific manner in which the index was calculated differed in each, and only one examines the index that is to be used as part of the BCP (or H/I ratio) test in a study that was still using statistical methods to find optimal cut points, i.e., a training study. So the *Oncotype DX* test, which has been validated in exactly the form given to patients on clinically homogeneous samples with clear treatment implications, is regarded as the index with the strongest claim to clinical validity. It is not yet as clear to which populations *MammaPrint* best applies, and how much incremental value it would have within those clinically homogeneous populations above various standard predictors. Since the number of validation studies for any of the tests is still relatively small, more remains to be learned about stability between different populations of the relationship between expression-based score and the absolute observed risk. Essentially nothing is known about how specific characteristics of these populations might affect test performance.

While the H/I ratio test shows some promise, it must be regarded as still being in a developmental phase; it cannot yet be considered fully validated. It was not clear whether samples were processed by Quest Diagnostics, which hold the current license. There are a number of intriguing biological insights and plausible mechanisms to support the rationale for the test, but its consistent value in well-defined clinical settings has not yet been firmly established.

Key Question 4. What is the clinical utility of these tests?

- a. To what degree do the results of these tests predict the response to chemotherapy, and what factors affect the generalizability of that prediction?

- b. What are the effects of using these two tests and the subsequent management options on the following outcomes: testing or treatment related psychological harms, testing or treatment related physical harms, disease recurrence, mortality, utilization of adjuvant therapy, and medical costs.
- c. What is known about the utilization of gene expression profiling in women diagnosed with breast cancer in the United States?
- d. What projections have been made in published analyses about the cost-effectiveness of using gene expression profiling in women diagnosed with breast cancer?

Few studies addressed the clinical utility of *Oncotype DX* recurrence score (RS) in predicting the benefits of adjuvant chemotherapy, although the probability of recurrence represents an upper bound on the degree of absolute benefit. One fairly strong retrospective study produced preliminary evidence that the RS has predictive power in assessing the benefit of chemotherapy usage in ER-positive, lymph node negative breast cancer patients. This study was embedded within a large, well conducted RCT (National Surgical Adjuvant Breast and Bowel Project (NSABP B-20)). Some patients from the tamoxifen-only arm of the trial were in the training data sets for the *Oncotype DX* assay development, and this could potentially translate into a somewhat enhanced estimate of the discriminatory effect of *Oncotype DX*, although it is unlikely to eliminate entirely the effect seen here. Other studies produced preliminary evidence that the RS from the *Oncotype DX* assay has predictive power in assessing the likelihood of pathologic complete response after pre-operative chemotherapy with various drugs and regimens, although very limited sets of patients have been used. One study produced preliminary evidence that the RS cannot predict pathologic complete response after primary chemotherapy in advanced breast cancer patients.

One study produced preliminary evidence that the knowledge of the RS from the *Oncotype DX* assay can have an impact on the clinical management of patients diagnosed with ER positive, lymph node negative, and early breast cancer. However, it did not report specifically what the patients (or doctors) were told or understood about their absolute risk of recurrence, and therefore was minimally informative as to the actual risk thresholds used by women and their treating physicians, or whether absolute risks even entered into the decision.

There were no studies that addressed the clinical utility of the MammaPrint or H/I ratio tests.

Three published studies have addressed economic outcomes associated with use of the breast cancer gene expression tests. One study reported that using the 21-gene RT-PCR assay to reclassify patients who were defined by 2005 National Comprehensive Cancer Network (NCCN) criteria as low risk (to intermediate or high risk) would lead to an average gain in survival per reclassified patient of 1.86 years. The associated cost-utility of using recurrence score testing for this cohort was \$31,452 per quality-adjusted life-year (QALY) gained. The analysis also reported that using the 21-gene RT-PCR assay to reclassify patients who were defined by 2005 NCCN criteria as high risk (to low risk) was cost saving. In a hypothetical population of 100 patients with characteristics similar to those of the NSABP B-14 participants, more than 90 percent of whom were NCCN-defined as high risk, using the 21-gene RT-PCR assay was expected to improve quality-adjusted survival by a mean of 8.6 years and reduce overall costs by about \$203,000. However, the EPC team had only moderate confidence in the results of this analysis because the study was sponsored in part by the manufacturer of the 21-gene RT-PCR assay and the authors did not provide sufficient information about methodological and structural uncertainties as well as other potential sources of bias such as the derivation of the utility

estimates. Furthermore, the 2007 NCCN guideline indicates that the use of chemotherapy in these patients is now considered optional, further diminishing the usefulness of these projections.

The second study reported that use of the 21-gene RT-PCR assay was associated with a gain of 0.97 QALYs and a cost-utility ratio of \$4432 per QALY compared with use of tamoxifen alone, and a gain of 1.71 QALYs with net cost savings when compared with the chemotherapy and tamoxifen combination. However, the EPC team had little confidence in the results of this analysis, which was supported in part by the manufacturer, because the study did not meet many of the standards that the team used for appraising the quality of the analysis.

The third study compared the cost-effectiveness of the Netherlands Cancer Institute gene expression profiling (GEP) assay (MammaPrint) to the U.S. National Institutes of Health (NIH) guidelines for identification of early breast cancer patients who would benefit from adjuvant chemotherapy. The GEP assay was projected to yield a poorer quality-adjusted survival than the NIH guidelines (9.68 vs. 10.08 QALYs) and lower total costs (\$29,754 vs. \$32,636). To improve quality-adjusted survival, the GEP assay would need to have a sensitivity of at least 95 percent for detecting high risk patients while also having a specificity of at least 51 percent. The EPC team had confidence in the results of this analysis because it met most of the standards for appraising the quality of an economic analysis.

Based on the appraisal of these three studies, the overall body of evidence on economic outcomes was inconclusive.

Limitations of the Report

The report included only English publications and was restricted to three gene expression tests.

Limitations of the Literature and Implications for Future Research

There are several issues that concern all of these tests.

1. While all of the tests exhibit a fair bit of risk discrimination (i.e., separating patients into different risk groups), the calibration of the estimates (i.e., how close the predicted risk is to the observed risk) in varying settings is still not as well established. Of greatest interest is the observed risk in the lowest risk groups, since the absolute level of this risk is critical for informed decisionmaking, and patients may forego chemotherapy on the basis of this information.
2. The manner in which the tests are best used—in combination with other prediction scores, as continuous scores, or as categorical predictors—has not been established. In addition, the current cut-points for designation of Low and High risks (with or without an intermediate category) are not clearly derived from decision-analytic criteria.
3. The incremental value of these tests is best assessed from cross-classification tables that show how many subjects are placed in different risk categories (corresponding to different clinical decisions) by the addition of the information from the test in comparison

or in addition to standard predictors. Such tables have been developed for *Oncotype DX*, but for only one set of risk thresholds, and some of the conventional guidelines used for those comparisons have since been updated.

4. In practice, pre-analytic issues related to sample preparation, transport and processing could cause the tests to perform differently in practice than in investigational contexts; continued monitoring of test procedures and performance will be important as they are used more widely.
5. The relevance of validation studies in past tamoxifen-treated populations for current populations treated with aromatase inhibitors needs further research.
6. Studies examining the use of the tests should provide women and physicians with quantitative risk information and report how this alters clinical decisionmaking. The manner in which this risk information is presented should also be studied.

Oncotype DX

1. The role of the RS in guiding treatment of HER-2 positive patients is unclear, as most of these patients were classified in the high RS group in the initial trials.
2. While awaiting the TAILORx results, the findings of the Paik 2006 study predicting treatment benefit need independent confirmation.

MammaPrint

1. The prognostic value of the 70-gene signature has been assessed in different populations facing different therapeutic choices. In the analysis by van de Vijver and colleagues, 130 of the 295 patients received adjuvant therapy in a non-randomized fashion. Patients in the original development cohort were not treated, and Buyse validated the marketed assay in untreated patients. It is not yet clear which are the optimal patient populations for the use of this test, exactly what its performance is in those populations, and how many of its predictions would result in different therapeutic decisions. Larger independent validation studies in therapeutically homogeneous groups would be very valuable.
2. There is no evidence for the degree to which this test predicts the benefit of adjuvant chemotherapy.

Breast Cancer Profile (H/I ratio) Test

1. The BCP test is not yet as well validated as either of the other tests, with most of the supporting studies examining slightly different ways of either performing (e.g., different reference standards) or calculating the index. More work needs to be done documenting the risk discrimination and risk calibration of the marketed test in clinically homogeneous populations, as well as its incremental value.
2. There is no evidence for the degree to which this test predicts the benefit of adjuvant chemotherapy.

In addition to the conclusions above, a series of other observations were made on the basis of what was learned in this investigation.

Assay Validation

In general, it is clear that validation studies need to deal with populations for whom the decision-making implications of various risk groupings are clear. For all tests except *Oncotype DX*, both validation and development studies have been on mixed populations, without sufficient sample sizes to stratify into large enough homogeneous groups to guide clinical decisionmaking. In addition, validation samples are often re-used by other investigators; the pool of such samples in the public domain needs to be greatly expanded.

Potential for Scale Problems

One problem that may be faced in the future is that of the consequences of an increase in demand for these tests. Whether the degree of accuracy seen in investigational settings can be maintained with increasing demands should be monitored by scientific or regulatory bodies.

Genetic Variability and Gene Expression

It is unknown whether gene expression profiles are more or less likely than more traditional biomarkers to be generalizable beyond the populations in which they were initially developed. Gene expression may reflect fundamental biological tumor features, and thus be relatively stable across ethnic groups. This speaks to the importance of validating these tests in populations with varying genetic background. Of particular interest will be the variation of the observed absolute risk in those populations, and its correlates.

The Need for Databases, Reproducibility, and Standards

Consideration should be given to the development of databases with complete data on each patient tested with these and future tests (absent identifiers). The data should include all the analyses performed, laboratory logs, the raw and processed data, and all the information about procedures and analyses that have been performed to produce a risk estimate from a tumor sample.

Where is the Field Going?

We can expect many new tests, as well as new uses for the assays that already exist. More genes might be added to the signatures, and in the particular case of *MammaPrint* this will be possible without changing the experimental procedures, since the array contains more genes than the ones that are incorporated in the 70-gene signature. In this regard, we might also expect other modifications: subsets of the current signatures might be proposed as alternatives to current clinical risk factors, or be proposed in different populations or for different purposes. For *Oncotype DX*, a natural evolution could be related to its use as an alternative to immunohistochemistry and/or pathology to evaluate tumor Grade, S-phase index, ER,

progesterone receptor, and HER2 expression, since such genes are part of the set included in the assay. Reporting of individual gene expression results may also prove useful.

“Comparative Effectiveness” Studies

As these tests mature and proliferate, an important question will be how they compare to each other, and whether there is value in their combination. In the therapeutic domain, this has been called “comparative effectiveness” research. Such research has traditionally been difficult to fund by government or by industry, because it may not hold out as much therapeutic promise as new discoveries, and because industry understandably is not anxious to fund head-to-head comparisons with competitive products. This same dynamic could easily take hold in the risk prediction arena, with a proliferation of licensed prediction indices without any clear notion of what new ones are contributing over previous tests. In this perspective, development of future expression-based predictors should account for direct contrasts with “established” methods.

Conclusion

The introduction of these gene-expression tests has ushered in a new era in which many conventional clinical markers and predictors may be seen merely as surrogates for more fundamental genetic and physiologic processes. The multidimensional nature of these predictors demands both large numbers of clinically homogeneous patients to be used in the validation process, and exceptional rigor and discipline in the validation process, all with an eye toward how the test will be used in a clinical decisionmaking context. Every study provides an opportunity to tweak a genetic signature, but we must find the right balance between speed of innovation and development of scientifically and clinically reliable tools. Going forward, it will be important to harness, if possible, as much genetic and clinical information on patients who undergo these tests to facilitate achieving each goal without unduly sacrificing the other.

Evidence Report

Chapter 1. Introduction

Breast Cancer

Breast cancer is the most commonly diagnosed cancer in women.¹ This tumor is currently the second leading cause of cancer-related deaths in women in the U.S., with approximately 178,000 new cases and 40,000 deaths expected among U.S. women in 2007.¹ Treatment for breast cancer usually involves surgery to remove the tumor and involved lymph nodes. Frequently, surgery is followed by radiation therapy (in case of breast conservation or in women with large tumors or many involved lymph nodes), endocrine therapy (for essentially all women with tumors that are estrogen receptor (ER)-positive (see Appendix A^a for a list of acronyms), and/or chemotherapy (for women having a high risk for a poor outcome, such as those with large tumors, involved lymph nodes, advanced disease, or inflammatory breast cancer). Chemotherapy administered in addition to surgery is called “adjuvant” chemotherapy. More than three-quarters of all patients are expected to survive with this multi-modality approach.

One major challenge in breast cancer treatment relates to the decision about whether or not to use adjuvant chemotherapy. Although adjuvant chemotherapy can reduce the annual odds of recurrence and death for many women with breast cancer, especially those with ER-negative tumors,² it has considerable adverse effects. Even though most women with early-stage breast cancer are advised to undergo chemotherapy, not all will benefit from it and some may remain free of disease recurrence at 10 years without it, especially those with small tumors and ER-positive disease. Decisionmaking protocols have been proposed with the intent of guiding clinicians involved in breast cancer treatment. Examples include the National Institutes of Health (NIH) Consensus Development criteria,^{3,4} the St. Gallen expert opinion criteria,⁵ the National Comprehensive Cancer Network (NCCN) guideline,⁶ and the computer-based algorithm Adjuvant! Online,^{7,8} which produces risk assessment and recommendations based on patient information, clinical data, tumor staging, and tumor characteristics (including age, menopausal status, comorbidity, tumor size, number of positive axillary nodes, and ER status). In addition, measurement of the human epidermal growth factor receptor 2 (HER-2) is now established as another predictive marker and has been incorporated into some of these indices,⁹ as it serves to identify candidates for adjuvant therapy with the monoclonal antibody trastuzumab (Herceptin[®]; Genentec, Inc., San Francisco, CA). Such patients may also be candidates for adjuvant treatment with other new agents such as the tyrosine kinase anti-HER-2 inhibitor lapatinib (Tykerb[®], GSK, PA) and the anti-vascular epithelial growth factor (VEGF) receptor antibody bevacizumab (Avastin[®]; Genentech), which are being studied in trials now in progress. With the proliferation of treatment advances in breast cancer, treatment decisions have become more complex, thereby increasing the demand for tests and predictive models that could help identify those patients most likely to benefit from specific therapies.

Breast cancer is increasingly understood as a broad umbrella label, with various tumor subtypes exhibiting different prognoses and different responses to the various treatment options available for use in the adjuvant setting. Evidence from large randomized trials, and systematic reviews, forms the basis of the various treatment algorithms and nomograms described above. These tools help caregivers determine the risk of recurrence and death and the chances of

^a Appendixes cited in this report are provided electronically at: <http://www.ahrq.gov/clinic/tp/brcgenetp.htm>

benefiting from a specific therapy within a tumor subtype (e.g., anti-estrogens alone for ER-positive disease, trastuzumab for HER-2-positive disease). Unfortunately, the predictive utility of these tools for an individual patient within a specific tumor subset is quite limited, and a large number of patients with ER-positive disease or HER-2-positive disease still experience tumor recurrence and die from their disease despite having received adjuvant anti-estrogen therapy or trastuzumab, respectively. Therefore, there is great interest in developing, testing, and validating strong predictive markers that can be used in daily clinical practice to accurately identify those patients most likely to benefit from specific therapy options such as chemotherapy, endocrine therapy, and anti-HER-2 therapy, alone or in combination.

Gene Expression Profiling

Gene expression profiling (see Glossary, Appendix B) is an emerging technology for identifying genes whose activity may be helpful in assessing disease prognosis and guiding therapy. Gene expression profiling examines the composition of cellular messenger ribonucleic acid (RNA) populations. The identity of the RNA transcripts (see Glossary, Appendix B) that make up these populations and the number of these transcripts in the cell provide information about the global activity of genes that give rise to them. The number of mRNA transcripts derived from a given gene is a measure of the “expression” of that gene. Given that messenger RNA (mRNA) molecules are translated into proteins, changes in mRNA levels are ultimately related to changes in the protein composition of the cells, and consequently to changes in the properties and functions of tissues and cells in the body. However, only 2 percent of the genome (see Glossary, Appendix B) is translated into proteins, and little is known about how the expression of this 2 percent is controlled. The key intermediate is the transcriptome (see Glossary, Appendix B), which is made up of all the individual transcripts produced by the cell (see Figure 1).

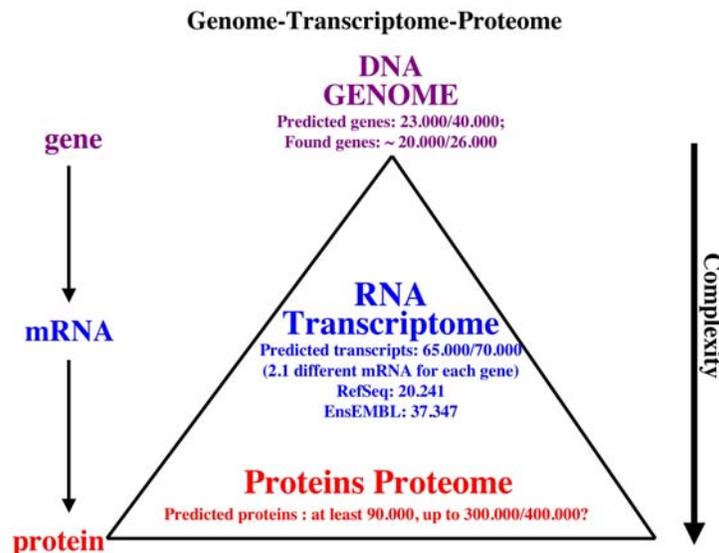


Figure 1: Increasing complexity of information from genome to transcriptome and proteome: gene expression profiling focuses on the analysis of the transcriptome.

Investigators have developed approaches to gene expression analysis that have led to substantial advances in our understanding of basic biology. Gene expression profiling has been applied to numerous mammalian tissues, as well as plants, yeast, and bacteria.¹⁰⁻¹⁴ These studies have examined the effects of treating cells with chemicals and the consequences of overexpression of regulatory factors in transected cells. Studies also have compared mutant strains with parental strains to delineate functional pathways. In cancer research, such investigation has been used to find gene expression changes in transformed cells and metastases, to identify diagnostic markers, and to classify tumors based on their gene expression profiles (see Glossary, Appendix B).¹⁵⁻¹⁸ The use of this approach for specific clinical problems, however, is relatively recent and poses several challenges related to the validity, reproducibility, and reliability required for use in diagnostic or predictive testing.

In recent years, gene expression profiling has been successfully used in breast cancer research. For instance, distinct subtypes of breast tumors (such as tumors expressing HER-2) have been identified as having distinctive gene expression profiles, representing diverse biologic entities associated with differences in clinical outcome.¹⁹⁻²³ Other investigators²⁴ have found gene expression signatures (see Glossary, Appendix B) associated with the ER and lymph node status of patients, thus identifying subgroups of patients with different clinical outcomes after therapy. From such studies, investigators have proposed a number of gene expression profiles that could be used to classify prognosis. In a case-control study from the Netherlands Cancer Institute (Amsterdam, the Netherlands), one such gene profile, consisting of 70 genes, was developed using archived frozen tissue from 78 young, node-negative women with breast cancer.²¹ In this study, tumors from patients who suffered rapid relapses after primary therapy had gene expression profiles that were quite distinct from those who remained disease-free. These gene expression profiles were then applied to a second validation set of 295 frozen tissue specimens collected from young women (including 61 patients from the previous cohort), yielding very similar results.²⁵ Indeed, it appeared that this 70-gene profile more accurately predicted outcomes than did the traditional clinical criteria. Results from these preliminary studies further suggested that gene expression profiling may provide a powerful tool for estimating prognosis and the likelihood of benefit from selected therapeutic agents.

Breast Cancer Assays on the Market

Three breast cancer gene expression profiling-based assays are now available in the U.S. These assays investigate the expression of specific panels of genes by measuring their RNA levels in breast cancer specimens using different techniques, real-time reverse transcription-polymerase chain reaction (RT-PCR)²⁶ (Glossary) and DNA microarrays²⁷ (see Glossary, Appendix B):

1. The *Oncotype DX*[™] Breast Cancer Assay (Genomic Health, Redwood City, CA) quantifies gene expression for 21 genes in breast cancer tissue by RT-PCR.²⁸ This test is intended to predict the likelihood of recurrence in women of all ages with newly diagnosed Stage I or II breast cancer, lymph node-negative and ER-positive, who will be treated with tamoxifen, an anti-estrogen agent.
2. The MammaPrint[®] Test is based on microarray technology, uses the 70-gene expression profile developed by van't Veer and colleagues,^{21,25} and is marketed by Agendia (Amsterdam, the Netherlands). This is a prognostic test for women 61 years of age or

younger with primary invasive breast cancer who are lymph node-negative and ER-positive or negative. The company voluntarily submitted this test to the U.S. Food and Drug Administration for approval under proposed new guidelines for such tests, and received such approval in February 2007. These guidelines were finalized in July 2007.

3. The Breast Cancer Profiling Test is based on the expression ratio of the two genes HOXB13 and IL17RB, and for this reason is also known as the H/I ratio test. The assay was developed by AvicaraDX and licensed to Quest Diagnostics, Inc. (Lyndhurst, NJ). This assay is based on RT-PCR and is offered to treatment-naïve women with ER-positive, lymph node-negative breast cancer.

All three tests have defined protocols for evaluating the tumor content of the specimens to be analyzed, preparing the RNA samples, normalizing the raw expression measurements, and computing summary indices which are related to patient prognosis. The characteristics of the assays, the gene panels used, and the procedures involved in the analysis are summarized in Table 1. Detailed descriptions of the genes can be found in Appendix C. These differences between tests must be taken into account in the evaluation of the available evidence about such tests. In the following section, we provide a brief description of the technologies that are used. A more detailed description is presented in Appendix D.

RT-PCR

RT-PCR is a molecular biology technique that combines reverse transcription with real-time PCR (see Glossary, Appendix B). This methodology allows the quantification of a defined RNA molecule. It is accomplished by reverse transcription of the specific RNA into its complementary DNA, followed by amplification of the resulting DNA using PCR. The quantification of the DNA produced after each round of amplification is accomplished by the use of fluorescent dyes that intercalate with double-stranded DNA, or by modified DNA oligonucleotide probes (see Glossary, Appendix B) that fluoresce when hybridized with complementary DNA.

In a PCR template, relative ratios of the product and reagent vary. At the beginning of the reaction, reagents are in excess, and template and product are present in low concentrations and do not compete with primer binding, so that the amplification proceeds at a constant, exponential rate. After this initial phase, the process enters a linear phase of amplification, and then in the late reaction cycles, the amplification reaches a plateau phase and no more product accumulates. To achieve accuracy and precision, it is necessary to collect quantitative data during the exponential phase of amplification, since in this phase the reaction is extremely reproducible. In RT-PCR, this process is automated, and measurements are made at each cycle. Finally, several implementations of this technique allow multiple DNA species to be measured in the same sample (multiplex PCR), since fluorescent dyes with different emission spectra may be attached to the different probes. Multiplex PCR allows internal controls to be co-amplified with the target transcripts (see Glossary, Appendix B) and permits allele discrimination in single-tube, homogeneous assays (Figure 2).

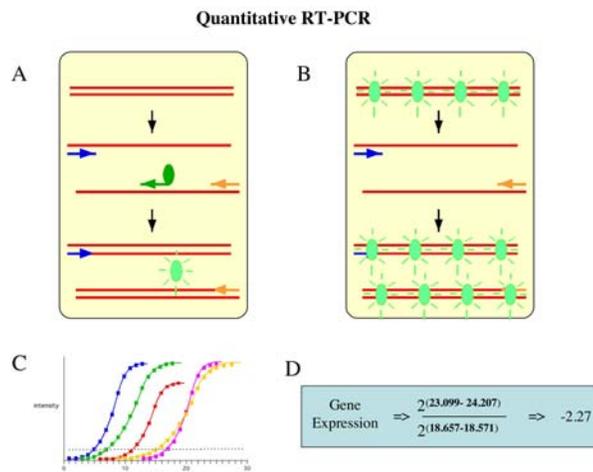


Figure 2: Quantitative RT-PCR. Panel A: PCR reaction using sets of quenched primers and probes. Panel B: binding of fluorescent probe molecules to double-stranded DNA. Panel C: fluorescence intensity curves for different dyes and samples: on the x-axis, the number of PCR cycle is shown, and on the y-axis, the corresponding fluorescence detected is indicated; the dashed line is used to calculate the cycle threshold for each sample. Panel D: computation of the relative levels of expression.

This technique is extremely sensitive. The development of novel chemistries and instrumentation platforms has led to widespread adoption of real-time RT-PCR as the method of choice for quantifying absolute changes in gene expression. Moreover, this technique has become the preferred method for validating results obtained from microarray analyses and other techniques that evaluate gene expression changes on a global scale.

Microarrays

The analysis of gene expression by microarray technology is based on the Watson-Crick pairing of complementary nucleic acid molecules. In this technique, a collection of DNA sequences, called probes (see Glossary, Appendix B), are “arrayed” on a miniaturized solid support (microarray) and used to detect the concentration of the corresponding complementary RNA sequences, called targets (see Glossary, Appendix B), present in a sample of interest. The advancements made in attaching or synthesizing nucleic acid sequences to solid supports and robotics have allowed investigators to miniaturize the scale of the reactions, and it is now possible to assess the expression of thousands of different genes in a single reaction.²⁹⁻³¹

In the basic microarray experiment, RNA harvested from the sample of interest is labeled with a fluorescent dye and hybridized to the microarray, then incubated in the presence of RNA from a different sample labeled with a different fluorescent dye. In this two-color experimental design, samples can be directly compared to one another or to a common reference RNA, and their relative expression levels can be quantified. After hybridization, gray-scale images corresponding to fluorescent signals are obtained by scanning the microarray with dedicated instruments, and the fluorescence intensity corresponding to each gene investigated is quantified by specific software. After normalization, the intensity of the hybridization signals can be compared to detect differential expression by using sophisticated computational and statistical techniques (Figure 3).

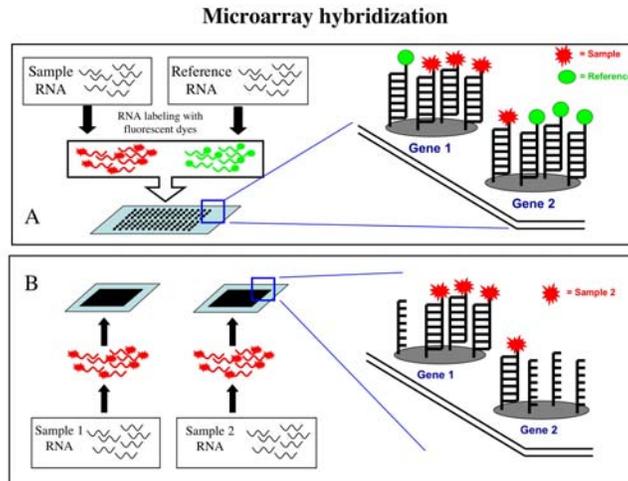


Figure 3: Schematic model for microarray hybridizations. Panel A: two-color scheme design. Panel B: single-color design.

Sources of Variability in Gene Expression Analysis

Gene expression analysis poses several general challenges that can affect the reproducibility and reliability of the measurements obtained. The control of such sources of variability is clearly a concern when such technologies are used to make decisions about the clinical management of patients. Given the complexity of the procedures used in this type of investigation, the sources of uncertainty are multiple, from the preparation of tissue specimens to the computational analysis used to quantify expression levels.

The first source of variability relates to the various types of specimens that can be used to prepare the RNA to be used in gene expression analysis, including tissue specimens obtained in vivo. In this case, the resulting RNA template will be a mixture of the RNA content of all the cells contained in the specimen, and the relative content of the different cell populations (malignant vs. normal) present in the specimen processed is a major source of variability in gene expression. For this reason, special care must be taken when tumors are sampled for gene expression analysis. In general, macro- or micro-dissection of the samples is performed to ensure that the specimens contain a sufficient percentage of cancer cells.

A second major source of variability is related to the protocols used to prepare the specimens, since several alternatives have been used in the field, including the use of formalin-fixed, paraffin-embedded (FFPE) tumor specimens or laser-captured, micro-dissected (see Glossary, Appendix B) specimens and fresh or snap-frozen samples. Other factors likely to affect RNA quality include storage time and the reagents, and particular batches used. Unlike DNA, RNA is very unstable. The degradation of RNA can be triggered by pH changes as well as by specific enzymes called ribonucleases (see Glossary, Appendix B) that are present in cells and that can remain active in the RNA preparation if the RNA isolation is not properly carried out.

Watson-Crick hybridization of complementary nucleic acid moieties is the fundamental principle that forms the basis of any gene expression analysis. For this reason, sequence selection and gene annotation (see Glossary, Appendix B) are among the most relevant factors that can contribute to variability in the analysis of gene expression.

As in any other laboratory investigation, the use of different platforms (see Glossary, Appendix B), protocols, and reagents can also affect the variability of the obtained measurements, and thus the reproducibility within and across laboratories. Indeed, numerous platforms exist to perform both RT-PCR and microarray-based gene expression analyses. Moreover, within each technique, the same procedure can be performed using different instruments, each with its own different operational characteristics and performance.

Finally, since gene expression measures are virtually never used as raw output but rather undergo sequential steps of mathematical transformation, another source of variability is data pre-processing and analysis. Moreover, the levels of gene expression can be further processed and combined according to complex algorithms to obtain composite summary measurements that are associated with the phenotypes investigated.

International standards have been developed to address the quality of microarray-based gene expression analysis, focusing on documentation of experimental design, details, and results (see MIAME in Glossary, Appendix B).³² Several publications also have addressed the levels of reproducibility across platforms and laboratories.^{33,34} Such efforts emphasize the importance of trying to control the many described sources of variability in gene expression analysis and of ensuring that the information derived from such analyses is specific and does not represent accidental associations.

Objectives of the Evidence Report

The overall purpose of this evidence report is to review and synthesize the available evidence concerning the analytic and clinical validity of breast cancer gene expression profiling in predicting disease recurrence, as well as its efficacy and effectiveness in improving chemotherapy choices and subsequent outcomes (clinical utility) in women newly diagnosed with early-stage breast cancer. The report was prepared by the Evidence-based Practice Center (EPC) at the Johns Hopkins University (JHU) Bloomberg School of Public Health in response to a task order issued by the Agency for Healthcare Research and Quality (AHRQ) on behalf of the Centers for Disease Control and Prevention (CDC) Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Project. The key questions we were charged with addressing in this evidence report were:

1. What is the direct evidence that gene expression profiling tests in women diagnosed with breast cancer (or any specific subset of this population) lead to improvement in outcomes?
2. What are the sources of and contributions to analytic validity in these gene expression-based prognostic estimators for women diagnosed with breast cancer?
3. What is the clinical validity of these tests in women diagnosed with breast cancer?
 - a. How well does this testing predict recurrence rates for breast cancer when compared to standard prognostic approaches? Specifically, how much do these tests add to currently known factors or combination indices that predict the probability of breast cancer recurrence (e.g., tumor type or stage, ER and HER-2 status)?
 - b. Are there any other factors, which may not be components of standard predictors of recurrence (e.g., race/ethnicity or adjuvant therapy), that affect the clinical validity of these tests and thereby the generalizability of the results to different populations?
4. What is the clinical utility of these tests?

- a. To what degree do the results of these tests predict the response to chemotherapy, and what factors affect the generalizability of that prediction?
- b. What are the effects of using these two tests and the subsequent management options on the following outcomes: testing- or treatment-related psychological harms, testing- or treatment-related physical harms, disease recurrence, mortality, utilization of adjuvant therapy, and medical costs?
- c. What is known about the utilization of gene expression profiling in women diagnosed with breast cancer in the United States?
- d. What projections have been made in published analyses about the cost-effectiveness of using gene expression profiling in women diagnosed with breast cancer?

This task is of particular relevance, since the National Cancer Institute (NCI) recently announced its sponsorship of a clinical trial to be conducted by The North American Breast Cancer Intergroup (TBCI) assessing individualized options for breast cancer treatment: the Trial Assigning Individualized Options for Treatment (TAILORx). In this trial, tumors of patients with ER-positive and lymph node-negative breast cancer (and who will be treated with tamoxifen) will be tested using the *Oncotype DX* assay, and patients will be divided into groups according to the recurrence scores derived from the use of the assay. Patients showing low recurrence scores will receive endocrine therapy alone, while patients with high recurrence scores will receive endocrine therapy and adjuvant chemotherapy. Patients with mid-range scores will receive endocrine therapy and be randomly assigned to chemotherapy or no chemotherapy. This trial is designed to evaluate the treatment implications of *Oncotype DX* results in a large representative patient population, focusing primarily on patients with intermediate recurrence scores. The trial will also allow for generation of new data on patients with recurrence scores near the ends of the spectrum. Patients at the low end of the recurrence score spectrum will be compared to a pre-specified target of 95 percent recurrence-free survival. It should be noted that the cutoff values used in the TAILORx trial are different than those delineated in other studies of *Oncotype DX*. The results of the TAILORx trial will not be available for some time (around 2013) and with growing interest in and use of these tests (particularly *Oncotype DX*) in the oncology community, this evidence review could have an impact on clinical practice in the interim.³⁵

A separate trial (MINDACT, or Microarray in Node-negative Disease may Avoid ChemoTherapy) has recently been activated by TRANSBIG (Translating molecular knowledge into early breast cancer management: building on the Breast International Group (BIG)), a research network of 39 institutions in 21 countries. The trial will compare two different ways of assessing the risk of cancer recurrence and making therapeutic decisions: a “traditional method” using Adjuvant! Online versus the MammaPrint assay. The rationale for this study is that many women who actually have “low risk” tumors are currently classified as “average” or “high risk” and therefore ultimately are recommended to receive adjuvant chemotherapy that ultimately may be of no benefit. The investigators estimate that 12-20 percent of women with early-stage breast cancer fall into this category.³⁶

Structured Approach to Assessment of the Questions

The EPC team used a structured approach to assess the evidence regarding the key questions listed above. The structured approach was based on the following questions:

1. What was tested? One fundamental concept is the distinction between the investigated gene expression signatures (see Glossary, Appendix B) and the actual gene expression-based tests. The gene “signature” is the collection of genes whose expression levels are measured in a given test, together with the algorithm that combines those levels into a prognostic index; akin to a test’s “recipe.” But just like a recipe can be implemented in subtly different ways with different results, this signature can be measured using a variety of technologies and procedures which may not be identical to those used in the actual marketed test being offered to patients. This distinction is important because clinicians’ decisions, patients’ choices, and the resulting benefits and harms will ultimately depend on the performance of marketed tests rather than on the more general gene expression signatures, although they typically track closely. Information about the signatures is highly relevant to the assessment of the marketed test, but is not identical.
2. What population was tested? This question required consideration of whether the study involved a representative sample of patients, from a clinical series or from a clinical trial subject to detailed eligibility criteria. This also required consideration of whether the population was clinically homogeneous enough for the implications of risk prediction to be clear and similar for every member of the study population (or for each subgroup). For example, predicting the relapse of patients on tamoxifen therapy may be different than predicting outcomes for untreated patients. The latter tests “intrinsic tumor aggressiveness,” which may not be the same as the factors that determine resistance to tamoxifen.
3. Was the study a developmental or validation study? Developmental studies were defined as the original reports in which new gene expression signatures were first described or in which previously developed gene expression signatures were first proposed to have a use different from the original use (e.g., the use on different subsets of patients with different purposes). Validation studies were defined as those that confirmed results in independent populations (with approximately the same characteristics as the population of the corresponding development study). If a developmental study, were appropriate statistical methods used to adjust for multiplicities, and was internal validation done? If a validation study, were all the test procedures, cutoffs, definitions, and measurements predefined?
4. Is it clear, from a clinical decisionmaking perspective, what is the incremental value of the test over and above standardized clinical predictors? It was not sufficient to simply insert clinical predictors into regression equations since this does not properly quantify the numerical consequences of decisions made with and without the new test.
5. Were the ways in which the tests had been evaluated optimal for clinical decisionmaking? This question required consideration of the choice of cutoffs, definition of categories, and combinations (or lack thereof) with other predictors.
6. What was the strength of the study design used to estimate clinical utility? Randomized controlled trials, with all samples taken concurrently, which could have taken place in the past, provide the strongest evidence of utility.
7. For studies of clinical utilization, what specific information was provided to patients and their physicians? Such studies are informative only if they are specific about the information that was given and how it informed decisionmaking.

Using this structured approach, the EPC team evaluated the evidence regarding the key questions of analytic validity, clinical validity, and clinical utility of each test, evaluated

separately. The EPC team then used the review of the evidence to formulate both test-specific and general conclusions.

Chapter 2. Methods

The CDC submitted a request for an evidence report on the “*Impact of Gene Expression Profiling Tests on Breast Cancer Outcomes*” to the AHRQ on behalf of the EGAPP. This evidence report will be used to inform the CDC’s Working Group as part of their work in formulating evidence-based recommendations. Our project consisted of recruiting technical experts, formulating and refining the specific questions, performing a comprehensive literature search, summarizing the state of the literature, constructing evidence tables, and submitting the evidence report for peer review.

Recruitment of Technical Experts and Peer Reviewers

At the beginning of the project, we assembled a core team of experts from JHU who had strong expertise in medical oncology, clinical trials, and biostatistics as well as a special interest in gene expression profiling tests. We also recruited external technical experts from diverse professional backgrounds, including academic, clinical, and corporate settings. The core team asked the technical experts and members of the EGAPP working group to give input regarding key steps of the process, including the selection and refinement of the questions to be examined. Peer reviewers were recruited from professional societies with an interest in breast cancer and gene expression profiling tests. Representatives from Agendia (MammaPrint[®]), Genomic Health, Inc. (Oncotype DX[™]), and Quest Diagnostics, Inc.[®] (BCP or H/I ratio) were also asked to review the report (see Appendix E^a).

Key Questions

The core team worked with the technical experts and representatives of the EGAPP and AHRQ to develop the Key Questions that are presented in the Specific Aims section of Chapter 1 (Introduction). The Key Questions apply to any gene expression profiling test, but they have been focused primarily on two gene expression profiling tests; Oncotype DX, and MammaPrint, because these are the tests that were expected to be commercially available in 2007. During the course of this review, the third gene expression profiling test, the Breast Cancer Profiling (BCP, or H/I ratio) Test (AviaraDX through Quest Diagnostics, Inc.) came to our attention. Although the BCP test was not included in our initial consideration of the Key Questions, we added studies regarding this test as an example of the types of gene expression profiling tests that are likely to be available in the coming years.

Literature Search Methods

Searching the literature involved identifying reference sources, formulating a search strategy for each source, and executing and documenting each search. For the searching of electronic databases we used medical subject heading (MeSH) terms that were relevant to breast cancer and

^a Appendixes cited in this report are provided electronically at: <http://www.ahrq.gov/clinic/tp/brcgenetp.htm>

gene expression profiling. We used a systematic approach for searching the literature to minimize the risk of bias in selecting articles for inclusion in the review. In this systematic approach, we were very specific about defining the eligibility criteria for inclusion in the review. The systematic approach was intended to help identify gaps in the published literature.

This strategy was used to identify all the relevant literature that applied to our Key Questions. The team specifically looked for articles that would provide information about the gene expression profiling tests identified in the Key Questions. We also looked for eligible studies by reviewing the references in eligible studies and pertinent reviews, by querying our experts, by contacting the manufacturers of the two tests, and by reviewing abstracts from relevant professional conferences.

Sources

Our comprehensive search plan included electronic and hand searching. On January 9, 2007, we ran searches of the MEDLINE[®] and EMBASE[®] databases, and on February 7, 2007, we searched the Cochrane database, including Cochrane Reviews and The Cochrane Central Register of Controlled Trials (CENTRAL), and CINAHL[®]. All searches were limited to articles published in 1990 or later. This cut-off year was established based on the introduction date of the MeSH heading “gene expression profiling,” 2000, and the introduction date of the MeSH heading “gene expression,” 1990. Also, test searches of earlier dates returned limited and irrelevant results.

“Gray” literature was searched following a protocol that was reviewed and approved by EGAPP and the technical expert panel:

1. Conference abstracts were reviewed using the same criteria as for journal articles but were only included if we felt we had a sufficient understanding of the underlying study and the data reported were critical enough to merit inclusion.
2. Web sites for the gene profiling tests included in this review, Agendia (MammaPrint[®]) and Genomic Health (Oncotype DX[™]), were searched for additional information not available in the peer-reviewed literature.
3. Agendia and Genomic Health, Inc. were contacted directly with requests for the following information:
 - a. A listing of articles that applied to the analytic validity or clinical utility of the gene profiling test,
 - b. Marketing materials on the gene profiling test, and
 - c. Any pertinent unpublished data.
4. We searched the Web site of the Food and Drug Administration (FDA) Center for Devices and Radiological Health for additional publicly available, unpublished information.³⁷⁻³⁹
5. A request was sent to the Center for Medical Technology Policy (CMTP) Gene Expression Profiling for Early Stage Breast Cancer Work Group to provide all background materials available on our study topic.

Search Terms and Strategies

Search strategies specific to each database were designed to enable the team to focus available resources on articles most likely to be relevant to the Key Questions. We developed a

core strategy for MEDLINE, accessed via PubMed, based on an analysis of the MeSH terms and text words of key articles identified a priori. The PubMed strategy formed the basis for the strategies developed for the other electronic databases (see Appendix F).

Organization and Tracking of the Literature Search

The results of the searches were downloaded into ProCite[®] version 5.0.3 (ISI ResearchSoft, Carlsbad, CA). Duplicate articles retrieved from the multiple databases were removed prior to initiating the review. We then reviewed the citations by scanning the titles, abstracts, and the full articles as described below (Figure 4).

Title Review

To efficiently identify citations that were obviously not relevant, paired reviewers first independently scanned the article titles. For a title to be eliminated at this level, both reviewers had to indicate that it was clearly ineligible (see Appendix G, Title Review Form).

Abstract Review

Inclusion and Exclusion Criteria

The abstract review phase was designed to identify articles that reported on the analytic validity, clinical validity, and/or clinical utility of the gene expression profile tests of interest. Abstracts were reviewed independently by two investigators and were excluded only if both investigators agreed that the article met one of the following exclusion criteria:

1. The study applied only to breast cancer biology;
2. The study did not involve *Oncotype DX* or *MammaPrint*,
3. The study did not involve original data or original data analysis;
4. The study did not involve women;
5. The study did not involve breast cancer patients;
6. The study was not in the English language; or
7. The study did not apply to the key questions.

We excluded letters to the editor and editorials when they did not present original data (usually in the form of electronic supplements in the case of letters). If a letter or editorial cited some original data, it generally was not sufficiently original for consideration in this report. As mentioned earlier, the initial scope of this project did not include the H/I ratio test, and thus this test was not identified on the abstract review form (Appendix G, Abstract Review Form).

Abstracts were promoted to the article review level if both reviewers agreed that the abstract could apply to one or more of the key questions. Differences of opinion regarding abstract eligibility were resolved through consensus adjudication.

Article Inclusion/Exclusion

Full articles selected for review during the abstract review phase underwent another independent review by paired investigators to determine whether they should be included in the full data abstraction. At this phase of review, investigators determined which of the Key Questions each article addressed (see Appendix G, Article Inclusion/Exclusion Form). If articles were deemed to have applicable information, they were included in the final data abstraction. Differences of opinion regarding article eligibility were resolved through consensus adjudication. A list of articles excluded at this level is included in Appendix H.

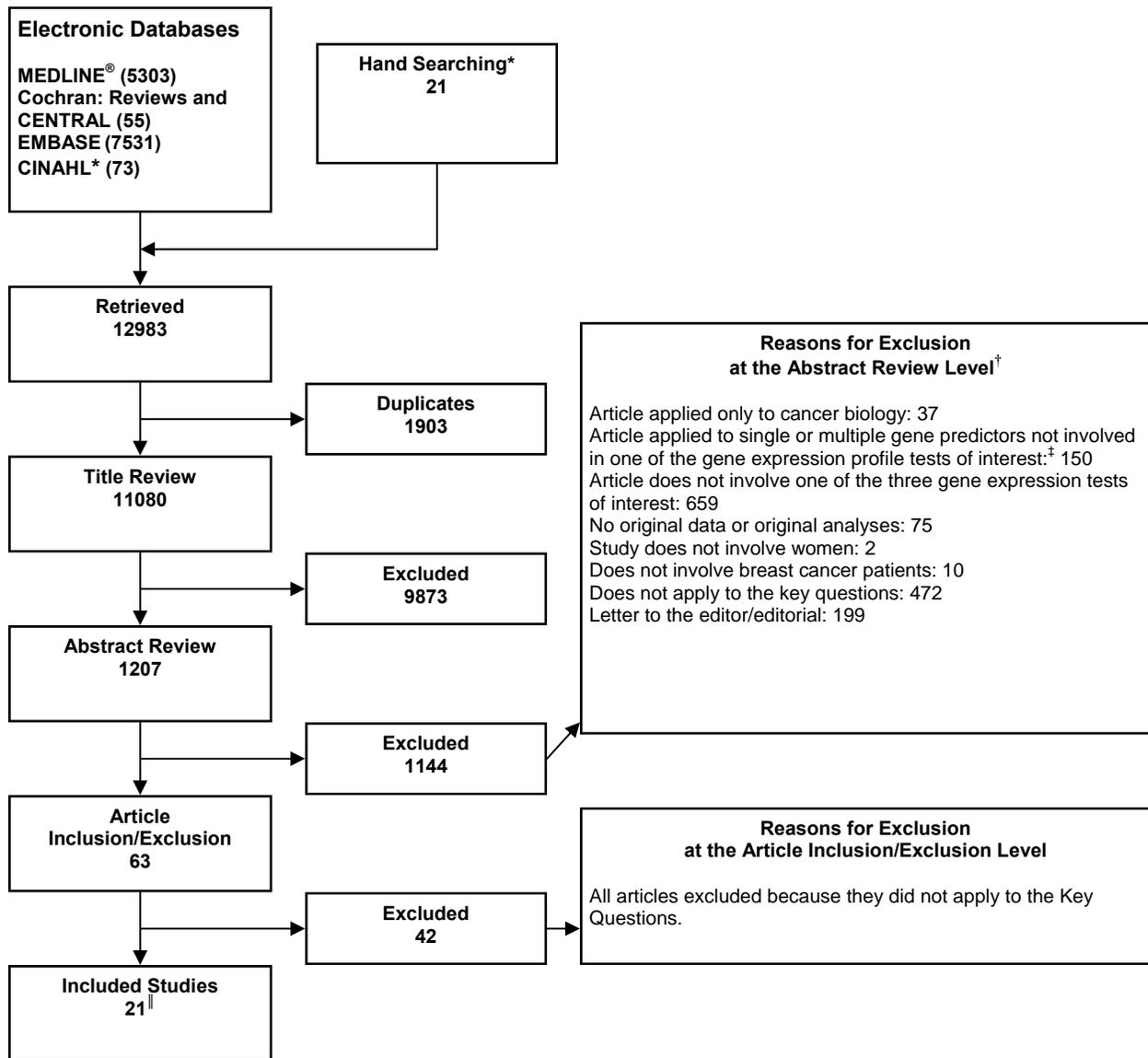


Figure 4. Summary of literature search and review process (number of articles).

* See Methods section for details

[†] Total is greater than 1144, reviewers were allowed to choose more than one reason for exclusion at this level.

[‡] Gene expression profile tests of interest: Oncotype DX (Genomic Health, Inc.), MammaPrint (Agendia), and The Two-gene Ratio (Quest Diagnostics, Inc.)

^{||}Total number of articles retrieved, 25. The Two-gene Ratio articles are not included in the body of the report but were pulled as articles of interest for comparison. One article applied to both the MammaPrint test and the Two-gene ratio test.

Data Abstraction

The purpose of the article review was to confirm the relevance of each article to the research questions and to collect evidence that addressed the questions. Articles eligible for full review had to address one or more of the Key Questions. Because of the heterogeneous nature of the applicable literature, we used a loosely structured approach for extracting data from the studies. Reviewers were given a standard matrix in which to enter data from each article (Appendix G, Data abstraction tables).

For all the data abstracted from the studies, we used a sequential review process. In this process, the primary reviewer completed all data abstraction forms. The second reviewer checked the first reviewer's data abstraction forms for completeness and accuracy. Reviewer pairs were formed to include personnel with both clinical and methodological expertise. Reviewers were not masked to the articles' authors, institutions, or journal.⁴⁰ In most instances, data were directly abstracted from the article. If possible, relevant data were also abstracted from the figures. A number of articles provided links to supplemental data, and these resources were used during the data abstraction process. Differences of opinion were resolved through consensus adjudication.

For all articles, reviewers extracted information on general study characteristics, such as study design, study participants, and sample size (see Appendix G, Data abstraction tables). Data abstracted regarding participants' characteristics were: information on intervention arms, age, menopausal status, race, diagnoses, methods of diagnosis, exclusion criteria, treatments, and treatment outcomes.

An analytic validity (Key Question 2) data abstraction matrix was developed by the team (see Appendix G, Data abstraction tables). Our data abstraction was designed to capture data in the following general areas: tumor specimens' processing validity, annotation validity; within- and across-laboratory validity; and validity associated with gene expression data preprocessing and analysis.

Studies addressing clinical validity (Key Question 3a, 3b) and utility (Key Question 4a, 4b, 4c) were approached in a similar manner (see Appendix G, Data abstraction tables). The free-form tables developed for these questions were designed to capture details regarding a study's context, the methods used to analyze the data collected, results of the study, and conclusions made by the study authors.

Only three articles addressed the cost-effectiveness of the gene expression profiling tests. Therefore, the reviewers did not use standardized data abstraction forms to abstract results from these studies. Instead, the reviewers extracted information directly into the table that is presented as Evidence Table 5. Please refer to the Philips, 2004⁴¹ article for a detailed explanation of why these domains and their sub-domains are important.

Quality Assessment

We used a synthesis of the general principles of the REporting recommendations for tumour MARKer prognostic studies (REMARK)⁴² and Standards for Reporting of Diagnostic Accuracy (STARD)⁴³ guidelines. The REMARK guidelines were developed to encourage transparent and relevant reporting of study design, preplanned hypotheses, patient and specimen characteristics, assay methods, and statistical analysis methods, in order to help others judge the usefulness of

the data presented.⁴² STARD was developed to improve the accuracy and completeness of studies reporting diagnostic accuracy, in order to allow readers to assess the potential for bias in a study and to evaluate the generalizability of the results⁴³ (Appendix G, Quality Assessment Matrix).

Because of the extreme variability of the articles included in this report, we did not systematically apply the general principles to them. The strengths and weaknesses of each study were also dependent on the question(s) to which it applied. These strengths and weaknesses are highlighted in the Results section and the Discussion.

The EPC team appraised economic analyses using published guidelines for good practice in decision-analytic modeling in health technology assessment (Phillips 2004). The appraisal took into consideration the domains of structure, data, and consistency (see Evidence Table 5 for details).

Data Synthesis

We created a set of detailed evidence tables containing all the information extracted from eligible studies and stratified the tables according to the gene expression profile test. The investigators reviewed the tables and eliminated items that were rarely reported. They then used the resulting versions of the evidence tables to prepare the text of the report and selected summary tables.

Data Entry and Quality Control

Initial data were abstracted by the investigators and entered directly into the data abstraction tables. Second reviewers were generally more experienced members of the research team, and one of their main priorities was to check the quality and consistency of the first reviewers' answers. In addition to the second reviewers checking the consistency and accuracy of the first reviewers, a senior investigator examined all reviews to identify problems with the data abstraction. If problems were recognized in a reviewer's data abstraction, the problems were discussed at a meeting with the reviewers. In addition, research assistants used a system of random data checks to assure data abstraction accuracy.

Grading of the Evidence

After reviewing the available evidence on the Key Questions, the core team concluded that it would be inappropriate to grade the overall body of evidence using any of the published schemes for grading evidence. None of the grading schemes fit the nature of the data in these studies about gene expression profiling tests. The team therefore decided that it was more appropriate to focus on the specific strengths and weaknesses of the studies on each Key Question.

Peer Review

Throughout the project, the core team sought feedback from the external technical experts and the EGAPP Working Group through ad hoc and formal requests for guidance. A

draft of the report was sent to the technical experts and peer reviewers, as well as to representatives of AHRQ, the CDC, the NIH, and the FDA. In response to the comments from the technical experts and peer reviewers, we revised the evidence report and prepared a summary of the comments and their disposition that was submitted to the AHRQ.

Chapter 3. Results

Key question 1. What is the direct evidence that gene expression profiling tests in women diagnosed with breast cancer, or any specific subset of this population, lead to improvement in outcomes?

In a study defined as providing direct evidence of improvement in outcomes, the use of the test in decisionmaking is compared to not using the test, with health outcomes as an endpoint, generally in the form of an RCT. There is currently no direct evidence that the investigated gene expression profiling tests lead to improvement in outcomes in any subset of women diagnosed with breast cancer. Two ongoing RCTs aim to provide almost direct evidence for *Oncotype DX*TM, and for *MammaPrint*[®]. These studies are described at the end of this chapter.

Key question 2. What are the sources of and contributions to analytic validity in these gene expression-based prognostic estimators for women diagnosed with breast cancer?

Analytical validity is usually assessed by determining how much observed measurements differ from expected values derived from a standard reference method. In the measurement of gene expression, however, universal standard reference RNAs and universally accepted, definitive methods of analysis are not available. Consequently, a definitive evaluation of the analytic *validity* of such type of test is difficult. It is more appropriate to focus instead on test *variability*. In clinical use, gene expression-based prognostic tests involve multiple steps with individual components that are difficult to separate. Ultimately, reproducibility of patient classification into clinically relevant risk groupings is what matters. From this perspective, the most important sources of variability are tumor sampling and handling, specimen preparation, and biologic variation within and between different samples of the same tumor. The analytic validity of expression-based tests can therefore be assessed by asking the following questions:

1. How reproducible is the test when applied repeatedly to the same patient, either by examining the same specimen, or a different specimen?
2. How reproducible is the test over time?
3. What are the factors that most affect the overall performance of the test?

Few existing studies directly address analytical issues involved with the assays, and additional information could only be collected from clinical studies. Overall, this evidence was heterogeneous, spanning technical aspects, reproducibility, the number of successfully performed assays, or the comparison of RNA and protein levels of individual genes. Table 1 describes the three assays; *Oncotype DX*, *MammaPrint*, and the H/I ratio (HOXB13 and IL17RB).

Oncotype DX™

Evidence about the analytic validity of Oncotype DX is available from two technical studies, Cronin et al, 2004,⁴⁴ and Cronin et al. 2007,⁴⁵) and from several clinical reports. Information about the overall success rate of the assay was documented in 9 studies (Chang, 2007,⁵⁵ Cobliegh, 2005,⁴⁷ Esteva, 2005,⁴⁸ Gianni, 2005,⁴⁹ Habel, 2006,⁵⁰ Mina, 2006,⁵¹ Oratz, in press,⁵² Paik, 2004,²⁸ and Paik, 2006⁵³). This success rate ranged from 78.9 percent to 98.9 percent, and only some of the studies provided detailed descriptions of the reasons for assay failure. Reported failures were mainly ascribed to an insufficient number of cancer cells in the specimens, to poor RNA quality, and in a few cases, to failure of the RT-PCR technique. A synopsis of this evidence is provided in Table 2.

Data on assay variability and reproducibility were available from 3 studies (Cronin, 2007,⁴⁵ Habel, 2006,⁵⁰ and Paik, 2004²⁸). These studies assessed the variability of repeated gene expression measurements using RNA from either the same or different FFPE blocks at repeated time points, and across different instruments and operators. Data reported in the study concerned the variability of individual genes in the assay as well as the RS reproducibility. Variability evidence was reported for 66 FFPE blocks from 22 distinct patients, and from repeated measurements of two aliquots of a pooled reference RNA. Overall, the standard deviation (SD) for the recurrence score was below 3 RS units, although the authors did not discuss the impact on risk stratification. This evidence is reported in Table 3.

Two studies addressed technical and operational aspects of analytic validity (Cronin, 2004,⁴⁴ and Cronin, 2007⁴⁵). The first study presented data about the development the assay procedures, comparing gene expression measurements between frozen tumor specimens and FFPE blocks. The optimization of the RT-PCR primers (see Glossary, Appendix B¹), and the normalization strategy were discussed. The second study addressed relevant analytic components of the assay, such as detection and quantification limits (limit of detection (LOD) and limit of quality (LOQ) respectively), amplification efficiency, linearity, dynamic range, accuracy, precision, and assay reproducibility. The available evidence is reported in Table 4.

Finally, eight studies (Chang, 2007,⁵⁵ Cobliegh, 2005,⁴⁷ Esteva, 2005,⁴⁸ Gianni, 2005,⁴⁹ Cronin, 2004,⁴⁴ Habel, 2006,⁵⁰ Mina, 2006,⁵¹ and Paik, 2004²⁸) compared gene expression measurements of specific individual genes (ER, progesterone receptor (PR), HER-2) to measurements of the corresponding proteins produced by those genes as obtained by other techniques, in particular immunohistochemistry (IHC). Such studies used various cycle thresholds (CTs) to define positivity for the genes (see Table 5). Overall agreement between RT-PCR and IHC proved generally good for ER (k statistics ranging from 0.80 to 1). In one study (Habel, 2006⁵⁰), agreement was low (0.49), although RT-PCR measurements were comparable to data available in the clinical records. In general, agreement for PR and HER-2 was moderate or poor. Such evidence is reported here for completeness (see Table 5), although it does not contain any relevant information about the assay as a whole.

Individual studies are briefly described below.

¹ Appendixes cited in this report are provided electronically at: <http://www.ahrq.gov/clinic/tp/brcgenetp.htm>

Cronin et al., 2004.⁴⁴ In this study, the authors discussed the primer (see Glossary, Appendix B) design optimization and expression level normalization necessary to obtain reliable RT-PCR measurements from archival FFPE samples, with the goal of establishing the reliability of their results with partially degraded RNA samples. The authors compared gene expression levels in 62 matched FFPE and frozen tissue specimens prepared from the same breast tumor. They showed that the relative expression profiles obtained from the two analyses were similar (correlation = 0.91, P value < 0.0001), although the magnitude of the measurements differed. They successfully corrected the differences using normalization based on the expression of five reference genes. Convincing evidence supporting the use of the implemented protocols in assessing gene expression levels from archival (i.e., formalin-fixed, paraffin-embedded) tumor specimens was shown.

The authors also analyzed several genes that were reported to show similar patterns in the literature²⁰ for co-expression,⁵⁴ and confirmed these correlations. Specifically, the expression of cytokeratin 5 and cytokeratin 17 ($r = 0.85$), LPL and RBP4 ($r = 0.84$), HER-2 and GRB7 ($r = 0.71$), ER1 and GATA3 ($r = 0.6$) were highly correlated.

Additionally, the authors compared RT-PCR measurements of ER, PR, and HER-2 (all components of *Oncotype DX*) expression to IHC analysis of corresponding protein levels, and to fluorescent *in situ* hybridization (FISH) analysis for HER-2 for a subset of 17 samples. The concordance among the different assays detecting the protein products of the genes and the relative RNA levels as measured by RT-PCR was high (94 percent, 84 percent, and 100 percent, respectively, see Table 5).

In summary, this study provided a foundation for the use of the *Oncotype DX* assay in archival tissue, although it did not contain data about the development of the RS (Appendix I, Evidence Tables 1, 2 and 3).

Paik et al., 2004.²⁸ In this clinical study, the authors reported data on the variability of the RS, and the overall success rate of the assay. The authors evaluated the reproducibility of the *Oncotype DX* assay within and between FFPE blocks from the same patient. The *Oncotype DX* assay was carried out on 5 serial sections from 6 different blocks from 2 distinct patients. Seventy nine blocks out of 754 were not analyzed due to insufficient tumor content, but RT-PCR was successful in 668 of the remaining 675 (98.9 percent) tissue blocks.

For the 16 genes considered in the RS, the SD of expression ranged from 0.07 to 0.21 expression units across serial sections from the same block. The within-block SD of the combined RS proved to be 0.72 RS units (with 95 percent CI: 0.55–1.04), while the within-patient SD, which included both among-block and within-block variation, proved to be 2.2 RS units. The impact of this variation on the risk stratification provided by the RS was not discussed in the paper. The difference between the low- and high-risk groups is 14 RS units, far larger than the standard deviations reported. Although ER, PR and HER-2 were also assessed by other techniques, the agreement of the measurement obtained by the different technologies was not reported.

In summary, this paper reported evidence about the fraction of tissue blocks that can be successfully typed by the *Oncotype DX* assay, as well as limited data about the reproducibility of the RS between different sections and FFPE blocks from the same patient. The impact of such variability on the risk stratification was not examined (Table 3, Appendix I, Evidence Tables 1, 2 and 3).

Esteva et al., 2005.⁴⁸ In this study, the authors evaluated the correlation of RS, both as a whole and broken into its components, with known standard prognostic markers in FFPE tumor specimens. Specifically, the relationship between RT-PCR and IHC for ER, PR, and HER-2 was examined. The concordance for PR status was poor (k of 0.48), high for ER (k = 0.81), and proved moderate for HER-2 (k = 0.60).

A logistic model using IHC HER-2 measurement as a quantal response indicated a significant ($P < 0.0001$) degree of correlation between IHC and RT-PCR. Sensitivity and specificity for HER-2 were also measured, using different RT-PCR cutoff points and positivity, and are reported in Table 5.

In summary, this paper reported evidence about the percentage of successfully-analyzed samples (67.7 percent, 149/220) in a large population from a single institution (M.D. Anderson Cancer Center) (Table 5, Appendix I, Evidence Tables 1, 2 and 3).

Cobleigh et al., 2005.⁴⁷ This study reports on the development of the 21-gene Recurrence Score assay (Oncotype DX). Duplicated gene expression measures were obtained by RT-PCR in archival FFPE tumor tissue blocks. An initial set of 192 genes (187 cancer-related and 5 controls) were analyzed and 16 additional candidate genes were added at a later time. Ninety-one point six percent (78/85) of samples were successfully analyzed

IHC-measured protein levels and RT-PCR mRNA levels for ER, PR, HER-2, and Ki-67/MIB-1 (a proliferation marker of cancer cells) were compared. The concordance was high for both ER (k = 0.83) and HER-2 (k = 0.67), somewhat lower for PR (k = 0.40), and poor for Ki-67 (k = 0.22). (Table 5, Appendix I, Evidence Tables 1, 2 and 3).

Gianni et al., 2005.⁴⁹ The authors of this paper evaluated the correlation of IHC-measured protein levels with RT-PCR mRNA measurements of ER and PR expression in tumors. The concordance was high for ER (k = 0.84; 95 percent CI, 0.71 to 0.96) and moderate for PR (k = 0.71; 95 percent CI, 0.56 to 0.86). This paper also reports preliminary evidence about the use of the Oncotype DX assay in fixed core biopsies from breast cancer patients. The percentage of successfully analyzed samples was 93.6 percent (89/95) (Table 5, Appendix I, Evidence Tables 1, 2 and 3).

Mina et al., 2006⁵¹ In this study, the authors evaluated the usefulness of FFPE core biopsies from a completed phase II trial in identifying genes that correlated with a response to primary chemotherapy. Out of the 70 patients enrolled in the study, 67 gave their consent, and specimens from 57 patients were available to perform gene expression analysis by RT-PCR. Out of these 57 patients, gene expression levels could be accurately measured in 45 patients. Failures were due either to low RNA yield (9 patients) or low tumor content in the biopsies (3 patients).

In this study the authors compared the expression levels of ER mRNA obtained by RT-PCR to ER protein expression as measured by IHC. Using a pre-defined cutoff of 6.5 CT, 64 percent of the 45 tumors were ER positive, while 36 percent were considered ER negative. ER expression by IHC correlated well with ER mRNA expression by RT-PCR (see Table 5), and only four of the 45 samples did not show agreement. The authors concluded that gene expression analysis on core biopsy samples was feasible. Data for PR, HER-2 and Ki-67 were not reported.

In summary, this paper reported preliminary evidence about the expression of some of the Oncotype DX assay genes in fixed core biopsies from breast cancer patients. The percentage of successfully analyzed samples was about 79 percent (45/57), raising concerns about the real feasibility in clinical settings (Table 5, Appendix I, Evidence Tables 1, 2 and 3).

Habel et al., 2006.⁵⁰ This study contains several results that are relevant for the overall analytic validity of the *Oncotype DX* assay. The authors cited two unpublished studies with data concerning the reproducibility of the RS. These studies analyzed, respectively, 60 blocks from a total of 20 distinct patients, and 49 core biopsies or resections from advanced breast cancer patients. In the first study the RS SD between different blocks from the same patient was 3.0 RS units, and less than 2.5 for 16 out of 20 patients. Similar results were claimed for the second study, although the actual data were not shown.

Finally, the authors compared the agreement of ER status, as obtained by RT-PCR, to the ER status reported in the medical records. A positive or negative classification was based on a CT cutoff point of 6.5. The RT-PCR failure rate was about 1 percent for specimens available after pathological review, and 7.9 percent of the samples were not assessable due to low tumor contents. In this study population, the concordance between RT-PCR and the medical chart information was only moderate ($k = 0.49$, 95 percent CI 0.41–0.56). In the multivariate models used in the following statistical analyses, the RT-PCR based ER status was used.

In summary, this paper reported a high percentage of successfully analyzed samples in a large population from a single institution and the reproducibility of the RS between different blocks from the same patient. The impact of such variability on the risk stratification was not addressed (Table 5, Appendix I, Evidence Tables 1, 2 and 3).

Paik et al., 2006.⁵³ In this clinical study the authors reported several results that can be used as indirect evidence for the overall analytic validity of the *Oncotype DX* assay. Particularly relevant, FFPE blocks with sufficient tumor content were available from 670 of the 2,299 eligible patients in the NSABP N-20 trial, and the RT-PCR assay was successful on 651 of the 670 patients (97.2 percent). (Appendix I, Evidence Tables 1, 2 and 3).

Cronin et al., 2007.⁴⁵ This study is the most extensive analysis to date of the analytic components of the *Oncotype DX* assay. Detection and quantification limits of the RT-PCR reactions, amplification efficiency, linearity, dynamic range, accuracy, precision, and assay reproducibility were investigated in serial dilution experiments, using a common RNA obtained by pooling 15 distinct RNA samples.

Detection and quantification limits proved to be well within the instrument's pre-specified CT unit limits for all the genes. Amplification efficiencies (100 percent efficiency means that the RT-PCR reaction products achieved perfect doubling) for the 16 cancer-related genes ranged from 75 percent to 112 percent, with an average of 96 percent, while the mean efficiency proved to be 88 percent for the reference genes, with a range from 75 percent to 101 percent.

Accuracy and precision studies were conducted at the target RNA concentration of 2 ng per assay well, which is what is used in the *Oncotype DX* assay. The mean percent bias from each gene target was -0.3 percent (ranging from -10 percent to 6 percent) for cancer-related genes, and 0.7 percent for reference genes (-1.5 percent to 3.3 percent), indicating 99 percent mean quantitative correctness at this assay condition. The CV averaged 5.7 percent for the cancer-related genes and 3.2 percent for reference genes. The implications of such variability for RS were not discussed.

Finally, individual gene and RS reproducibility were measured by performing repeated analyses across multiple days, operators, RT-PCR plates, RT-PCR instruments, and liquid-handling robots. Two operators obtained replicate CT measurements on two aliquots of a single RNA sample over the course of five days with three real time PCR instruments (7900HT instruments) and two liquid-handling robots. The study design allowed the estimation of all main effects, including operator, RT-PCR instrument, and liquid-handling robot. Total SD in CT

measurements varied from 0.06 to 0.15 CT units across the 21 genes, and the upper bounds on 2-sided 95 percent confidence intervals for the CV were all within 10 percent. The authors reported that a maximum SD of 0.15 at a CT of 30 translates into a CV of 0.5 percent, allowing a 15 percent change in gene expression to be distinguished. The day-to-day SD for all 21 genes ranged from 0 to 0.055, the between-plate SD ranged from 0 to 0.09, while the within-plate SD ranged from 0.057 to 0.147. The standard deviation for the overall RS (total and within-plate) was 0.8 RS unit. The largest differences between operators, as well as between liquid handling robots and 7900HT instruments, were 0.5 CT units for each of the 21 *Oncotype DX* genes, while SD and CV for the RS were not reported.

In summary, this study presented extensively detailed results about several relevant analytic components of the assay (Table 4, Appendix I, Evidence Tables 1, 2 and 3).

Chang et al., 2007.⁵⁵ This clinical study reported several results that can be used as indirect evidence for the overall analytic validity of the *Oncotype DX* assay. Ninety-seven FFPE blocks from core biopsies were analyzed by the standard assay protocols, and the percentage of successfully analyzed samples was 82.4 percent.

In summary, this paper provides preliminary evidence about the use of the *Oncotype DX* assay in fixed core biopsies from breast cancer patients (Table 2, Appendix I, Evidence Tables 1, 2 and 3).

Oratz et al., in press.⁵⁶ This clinical study evaluated the impact of the *Oncotype DX* assay on clinical management, and also provided indirect evidence for the assay's overall analytic validity. Seventy-four FFPE blocks were analyzed by the standard assay protocols, and the percentage of successfully analyzed samples was 97.3 percent. No explicit eligibility criteria were used. The samples were included based on the request for analysis from the patient's clinician.

In summary this paper contains evidence about the use of the *Oncotype DX* assay on FFPE blocks from breast cancer patients (Table 2, Appendix I, Evidence Tables 1, 2 and 3).

MammaPrint®

Analytic validity and variability evidence for MammaPrint was available from two technical studies (Ach, 2007,⁵⁷ and Glas, 2006⁵⁸) and information on the overall success rate of the assay was documented in just one study, Buyse, 2006⁵⁹ (80.9 percent).

Data about variability and reproducibility were obtained in these studies using repeated gene expression measurements over time, within and across individual microarrays, across different laboratories, protocols instruments, and operators (see Tables 6, 7, 8). No comparisons were made between expression measurements of individual genes and their corresponding protein level by IHC.

The following is a brief description of each study.

Glas et al., 2006.⁵⁸ In this study the authors reported a summary of the results obtained during the development of the commercially marketed version of the 70-gene prognostic signature,^{21,25} the expression array-based test known as MammaPrint. The authors evaluated and compared both technical aspects and the clinical validity of the assay using the originally published data (see Key Question 3).

MammaPrint uses a microarray accounting for 1,900 features (individual microarray locations where the probes are positioned), containing each of the 70 genes in the signature spotted in triplicates. In this paper the authors re-analyzed the data from the original series^{21,25}

using the new array, a dye-swap hybridization design, a different reference RNA and a different approach to computing gene expression levels. Triplicate measurements were obtained for each gene of the 70-gene signature and summarized by an error-weighted average, rather than the approach proposed by Hughes et al., 2000,⁶⁰ which was used in the original studies.

The results obtained with the new signature were comparable to the original results. Briefly, MammaPrint proved reproducible on the original development series²¹ (Pearson's correlation coefficient = 0.92 P value < 0.0001), and in a subset of the van de Vijver cohort²⁵ (Pearson's correlation coefficient of 145/151 lymph node-negative patients = 0.88, P value < 0.0001). The replication of the experiment within patients and along time suggested high reproducibility as well. In particular, the Pearson's correlation coefficient on 49 patients analyzed twice was 0.995, and no significant variability within individuals was found by an analysis of variance (ANOVA) for the 70-gene signature P value = 0.96).

Risk classification by MammaPrint is obtained by measuring the cosine correlation of individual patients' gene expression profiles to the mean gene expression profile obtained in the van't Veer²¹ series. The variability of such correlation was measured by repeated analysis of 3 patients over time and showed very small SDs (0,028, 0,028 and 0.027 respectively).

In summary, this study reported detailed data about the development of the MammaPrint assays as it is offered in clinical settings, as well as data about the reproducibility of the assay within a single laboratory (Tables 7 and 8, Appendix I, Evidence Tables 6, 7 and 8).

Buyse et al., 2006.⁵⁹ In this clinical study the authors reported several results that can be used as indirect preliminary evidence for the overall analytic validity of the MammaPrint® assay. Fresh frozen blocks from primary breast cancer patients collected in 5 distinct institutions were shipped for analysis to Agendia, and the percentage of successfully analyzed samples was 80.9 percent (326/403 patients) (Appendix I, Evidence Tables 6, 7 and 8).

Ach et al., 2007.⁵⁷ The inter-laboratory reproducibility of the MammaPrint assay was assessed in this paper. Results for the same set of four patients were obtained at three different sites and compared in order to assess the variation resulting from several important phases of analysis, including RNA amplification and labeling, hybridization and wash, and slide scanning. The same input RNA was used for all experiments.

In the first phase of the analysis, two laboratories, one in Amsterdam and one in California, amplified and labeled the RNA samples, then exchanged aliquots of the templates. Hybridization and slide scanning were performed at both locations and the scanned slides were then exchanged for re-analysis by the other laboratory. The same lot of labeling kits and microarrays were used at both sites. Technical replication variability was assessed by analyzing two separate slides in two different days. This experimental design allowed examination of both intra- and inter-laboratory variation.

The Pearson correlation coefficient across all technical replicates for all tumors analyzed proved to be above 0.983, indicating that the signals from replicate hybridizations correlated extremely well for genes expressed at all the measured intensity levels.

The reproducibility of laboratory scanning procedures was evaluated by scanning each of the 16 microarray slides at both sites. Signals for green fluorescent dye proved extremely reproducible, irrespective to the site of first hybridization and scan (Pearson correlation coefficient > 0.995, slope = 0.97), while signals for the red dye correlated less well and were always lower on the rescanned slide. The correlation of the 70-gene expression profile to the previously developed⁵⁹ mean signature⁵⁸ was computed for each dye-swapped pair of arrays and ANOVA was used to evaluate the variability by hybridization site, labeling site, and

hybridization day. No significant differences were found between hybridization sites, or hybridization days (regardless of site), but two tumors showed a statistically significant difference (P value <0.05) between labeling sites. Variability due to the RNA labeling site was further confirmed for expression measurements of individual genes of the 70-gene expression profile, as well as on the 182 most highly expressed genes.

In the second phase of the study, the assay performance was evaluated by a third laboratory in Paris, France, using a different batch of arrays, reagents, and labeling kits, on the same four tumor RNAs, several months after the initial comparison. The 70-gene signature correlation values for each of the four tumors were compared by ANOVA analysis, and significant differences were found for two of the tumors, when stratified by labeling site (P values of 0.0004 and 0.01 respectively), whereas one tumor proved to be significantly different (P value, 0.016) by hybridization site. The authors predicted, but did not provide supporting data, that if variations in the washing protocols were introduced between laboratories, significant discrepancies in the 70-gene signature results would emerge. They concluded that while some sources of variation have measurable influence on individual microarray measurements, the overall impact on the 70-gene signature is low.

In summary, this study thoroughly investigated factors that could affect the reproducibility of the 70-gene signature within and across different laboratories. RNA labeling proved to be the largest contributor to inter-laboratory variation, but the authors did not address the impact of such factors on the classification of individual patients into different risk groups. The data (although from only four distinct patients) implies that results from MammaPrint testing cannot be compared across laboratories and that the test must be centralized (Tables 7 and 8, Appendix I, Evidence Tables 6, 7 and 8).

H/I Ratio

None of the studies reviewed here explicitly referred to the marketed H/I ratio (BCP assay). However, one publication described the analytic procedures involved with such test, Ma, 2006.⁶¹ The rest of the available analytic validity and variability evidence was specific to the way in which the two-gene ratio profile was computed in each clinical study, and did not contain direct information about the marketed test.

Three studies (Goetz 2006,⁶² Jerevall 2007,⁶³ Ma 2006⁶¹) reported the overall success rate of the analyses, one report, Jerevall 2007,⁶³ assessed the reproducibility between two different institutions, one assessed the correlation between RT-PCR and microarray based gene expression measurements for the two genes (HOXB13 and IL17RB), and one, Ma 2004,⁶⁴ study compared ER status by RT-PCR and IHC (see Tables 9, 10, and 11). No comparisons were made between expression measurements of HOXB13 and IL17RB transcripts and the corresponding proteins by IHC. For completeness, a brief description of individual studies follows.

Ma et al., 2004.⁶⁴ In this study the authors developed the HOXB13/IL17BR two-gene ratio signature. They identified differentially expressed genes associated with breast cancer recurrence in patients who were treated with tamoxifen, using gene expression arrays on whole mount as well as on laser micro-dissected (LMC) specimens. From a total of 5,475 genes selected because of their high variability across tumors, three differentially expressed genes proved to be common between the two analyses (macro-dissected specimens vs. LCM). These genes were HOXB13 (identified twice as AI700363 and BC007092), the 17B receptor IL17BR (AF208111), and EST AI240933.

HOXB13 was found to be over-expressed in tamoxifen recurrence cases, whereas IL17BR and AI240933 were over-expressed in tamoxifen non-recurrence cases. The authors confirmed relative gene expression by RT-PCR microarray analysis on 59 out of the 60 original patients. The Pearson correlation coefficient between array and RT-PCR results was 0.83 for HOXB13, and $r = 0.93$ for IL17BR. The RT-PCR-derived HOXB13/IL17BR ratios also highly correlated with its microarray-derived counterpart (0.83). The authors also evaluated by RT-PCR 20 additional ER-positive early-stage primary breast tumors from women treated with adjuvant tamoxifen monotherapy between 1991 and 2000. These were used as a validation set (see Key Question 3).

In summary, this study provides a foundation for the use of the H/I ratio signature in LMC FFPE specimens (Table 10, Appendix I, Evidence Tables 10, 11 and 12).

Ma et al., 2006.⁶¹ The authors developed the two-gene index concept in this study, based on the two-gene ratio they originally published in Ma et al, 2004.⁶⁴ New RT-PCR primers/probes for HOXB13 and IL17BR were used, and four reference genes were introduced for normalization. Total RNA was isolated from two 7-micrometer thick tissue sections for each sample, reverse transcribed into cDNA using a pool of gene-specific primers, and quantitated by TaqMan RT-PCR in duplicate in a 384-well plate. For each sample, CT values for the four reference genes were averaged and the relative expression level of each target gene was expressed as the difference from mean reference CT after Z-transformation. This resulting value is no longer a simple ratio, and is thus referred to as the two-gene index.

RNA for this study was prepared from cancer cells isolated by LCM from FFPE tissue microarray sections (see Glossary, Appendix B) of originally frozen tumor specimens. From 870 patients, 98.0 percent of samples were successfully processed (Table 9).

In this study the authors evaluated the concordance between ER and PR protein levels assessed by IHC and the corresponding gene expression measured by RT-PCR. Since the distributions were found to be bimodal for both genes, the midpoints between the two populations were used as cutoff points (2.5 CT for ER and 5.9 for PR). Both the ER (91 percent concordance; kappa = 0.83; P value = .0001), and PR (85 percent concordance; kappa = 0.70; P value = .0001) status proved to be highly concordant. According to the authors, this confirms the significant correlations between mRNA and protein levels for ER and PR and provided validation of their gene expression analysis.

In summary, this clinical study, in which the HOXB13-to-IL17BR index was developed, represents the foundation for using the two-gene ratio signature in tissue microarray FFPE specimens analysis (Table 11, Appendix I, Evidence Tables 10, 11 and 12).

Goetz et al., 2006.⁶² In this clinical study, FFPE tumors samples from 206 of 211 primary breast cancer patients were successfully processed by laser micro-dissection (LMC) prior to total RNA preparation. This study provides generic evidence about the analytic validity of the two-gene signature in primary breast cancer patients, as computed from LMC processed FFPE blocks (Table 9, Appendix I, Evidence Tables 10, 11 and 12).

Jerevall et al., 2007.⁶³ This paper quantified expression of HOXB13 and IL17BR (normalized to beta-actin) by RT-PCR in fresh frozen specimens from two distinct institutions in Sweden. RT-PCR reactions at the two institutions were performed using the same sets of primers and fluorescent probes, and two distinct instruments. Ninety-six percent of the 373 samples were successfully analyzed.

In summary, good reproducibility of the measurement between institutions was documented for each individual gene and the ratio (Pearson's correlation coefficient = 0.99, P value < 0.001) (Table 10, Appendix I, Evidence Tables 10, 11 and 12).

Key Question 3. What is the clinical validity of gene expression profiling tests in women diagnosed with breast cancer?

A synopsis of the clinical validity evidence presented in the following section is reported in Table 12.

Oncotype DX

Paik et al., 2004.²⁸ This study was the first to validate the prognostic validity of *Oncotype DX* in a population independent from that used to develop the test. The population consisted of a sample of 668 (out of 2617) lymph node-negative, ER positive breast cancer patients from the tamoxifen-treated arm of the National Surgical Adjuvant Breast and Bowel Project (NSABP) Trial B-14. This 668-patient subset had enough analyzable tissue in paraffin blocks to be evaluated using the *Oncotype DX* assay, and was reported to be similar in baseline characteristics to the overall sample. A more complete sample was impossible because of sample unavailability or processing problems. In this study, the overall 10-year distant recurrence rate was 15 percent and the RS was significantly correlated with disease-free survival and overall survival ($P < 0.001$ for both). The authors reported that RS alone was a better predictor of the distant recurrence risk than traditionally used predictors. In a multivariate model including age, tumor size grade, ER, PR, and HER, the RS Hazard Ratio was 2.81 (95 percent CI, 1.70–4.64, $P < 0.001$, per 50 unit increase). Forty-four patients out of the 109 with small tumors (diameter less than 1 cm), were classified using *Oncotype DX* into the intermediate or high risk groups (Table 12, Appendix I, Evidence Tables 1, 2 and 4).

Esteva et al., 2005.⁴⁸ In this study the *Oncotype DX* assay was evaluated in a population of 149 patients treated at the MD Anderson Cancer Center between 1978 and 1995. These patients had been diagnosed with node-negative breast cancer and did not receive tamoxifen or chemotherapy, and had a median 18 year followup. The number of recurrences was not reported, and this study failed to find correlation between RS and distant breast cancer recurrence. ER, PR, and HER-2 showed no prognostic value, and well-differentiated tumors were correlated with worse survival than higher grade tumors, the reverse of expected. The population was unusual in that it received no treatment, and was different from the one used by Paik et al.²⁸ (Table 12, Appendix I, Evidence Tables 1, 2, and 4).

Cobleigh et al., 2005.⁴⁷ This report is the only study among the three used to develop the 21-gene Recurrence Score assay (*Oncotype DX*) to be published in a peer-reviewed journal. Seventy-eight breast cancer patients with more than 10 positive nodes from Rush University Cancer Center were studied, and 55 had recurred. Two hundred and fifty-five candidate genes were amplified with RT-PCR from FFPE tumor tissue obtained as long as 24 years ago. Twenty-two genes were significantly correlated with distant recurrence-free survival (DRFS) (unadjusted P value < 0.05). An RS was developed using these genes which very strongly predicted disease-free survival, but as this was training and not validation data, it has minimal evidential value in

assessing *Oncotype DX* predictive properties (Table 12, Appendix I, Evidence Tables 1, 2 and 4).

Habel et al., 2006.⁵⁰ The *Oncotype DX* assay was used to assess the risk of breast cancer-specific mortality among women in a large case-control study population derived from fourteen Northern California Kaiser community hospitals with ER positive, node-negative breast cancer.

There were a total of 4,964 eligible patients, 220 had died and 570 were living controls. All were younger than 75 years old, diagnosed between 1985 and 1994, and had not been treated with adjuvant chemotherapy. For ER positive tamoxifen-treated patients, RS risk groups (as defined by pre-specified thresholds chosen by the test developers) showed similar 10-year risks of death from breast cancer (3 percent, 12 percent, and 27 percent respectively for low, intermediate, and high risk, groups) as Paik²⁸ reported for the NSABP B-14 patients. Multivariate analysis showed that RS and tumor size were significant and independent risk predictors of breast cancer death in both ER positive, tamoxifen-treated (hazard ratio per 50 units = 7.6, $P < 0.001$) and untreated patients (RS hazard ratio per 50 units = 4.1, $P < 0.001$). Tamoxifen-treated patients were shown to have a higher risk of death, and tumor grade proved to be a significant, independent predictor as well. The RS score showed some prognostic value in ER negative patients, although this group was too small to perform a reliable analysis.

ER status was missing from the medical record for a substantial proportion of patients in this study, and therefore ER status based on gene expression was used in the analysis. Cases and controls were matched with respect to tamoxifen treatment, so it was not possible to assess whether the RS was able to identify patients who are likely to respond to tamoxifen therapy. The performance of the *Oncotype DX* assay RS was not compared to standard risk stratification methods (e.g., St. Gallen, NIH criteria, or Adjuvant! Online) (Table 12, Appendix I, Evidence Tables 1, 2, and 4).

Paik et al., 2004,⁶⁵ **Bryant 2005,**⁶⁶ **and Hornberger et al., 2005.**⁶⁷ These posters showed the cross-classified risk predictions of the *Oncotype DX* assays compared to the risk stratifications using the 2004 NCCN and 2003 St. Gallen criteria, with the observed 10 year risks of relapse in the cross-classified strata. NCCN guidelines have since been modified, and the St. Gallen criteria did not account for HER-2. Patients came from the Paik⁶⁵ NSABP-14 validation cohort, N=668. Using the 2004 NCCN guidelines, the study indicated that of the 92 percent who were in the high-risk NCCN category, about half were reclassified as low-risk by RS, with a 10-year relapse risk of 7 percent (95 percent CI, 4-11 percent), which is similar to the risk observed in the low risk RS group, without the NCCN information⁶⁵. Finally, against the Adjuvant Online criteria, roughly 40 percent of those assessed to be at high risk (22 percent relapsed) were reclassified as having an 8 percent risk if they had a low RS score. These data, demonstrate that optimal predictions may come from a combination of expression predictors and standardized indices, although the latter contribute less than the RS to the risk estimate (Tables 13, 14, and 15).

MammaPrint

A synopsis of the clinical validity evidence presented in the following section is reported in Table 16. In the following section we will be distinguishing between MammaPrint, the marketed assay, and the gene expression profile which is the 70-gene signature originally published by van't Veer et al., in 2002.²¹

van't Veer et al., 2002.²¹ This study reported the development data for the 70-gene panel that is the basis for the MammaPrint test. A gene expression array containing 25,000 features was used to select genes associated with metastases-free survival at 5 years from surgery in 78 node negative patients, including 34 patients who recurred at 5 years and 44 who had not. Using the development of metastasis within 5 years as the first relapse event, 65 out of the 78 patients were correctly classified into good and poor prognosis groups by the 70-gene signature. Among the 13 misclassified patients, 5 patients with poor prognosis were in the good prognosis group, while 8 patients with good prognosis were classified in the poor prognosis group. Seventeen of 19 were correctly classified in the validation set.

The odds ratio (OR) to develop metastases within 5 years was 28, (95 percent CI, 7-107), while after leave-one-out cross-validation it was 15 (95 percent CI, 4-56). Using univariate analysis, the 70-gene signature performed better than tumor grade, size, patient age (less than 40years), ER status, and angioinvasion. Using multivariate analysis, the 70-gene signature was an independent predictor of metastases within 5 years, OR = 18 (95 percent CI, 3.3-94) (Tables 16 and 17, Appendix I Evidence Tables 6, 7 and 9).

van de Vijver et al., 2002.²⁵ This was the first major validation of the 70-gene signature as reported in van't Veer 2002 using the same protocol and approach. Banked tumor specimens from the Netherlands Cancer Institute were used from a consecutive series of 295 women with breast cancer, with a mix of lymph node positivity, ER status, chemotherapy, and tamoxifen treatment. Time to metastases, as well as overall survival (OS) were used as primary end points in survival models, and 61 patients in this cohort had been in van't Veer's²¹ original 78 patient training set.

Patients were young (less than 52 years) with small tumors (less than 5 cm). The 70-gene signature was shown to be associated with grade, size and ER positivity, with almost all of ER positive patients falling into the good prognosis category. Those with "good prognosis" 70-gene expression signatures had dramatically better 5-year (95 percent vs. 61 percent) and 10-year (85 percent vs. 51 percent) DRFS and OS (95 percent vs. 55 percent at 10 years) than the "poor prognosis" group. Multivariate analysis showed that the prognosis group, tumor size, and adjuvant chemotherapy were the strongest predictors of distant metastases. The "poor prognosis" signature had the largest hazard ratio = 4.6 (95 percent CI, 2.3-9.2). Analyses excluding the 61 previously-included patients produced similar results. Fourteen of the 115 "good signature" patients experienced a recurrence by 10 years, demonstrating that the "good prognosis" group may not be at low enough long-term risk to justify forgoing chemotherapy when the 70-gene signature is used alone.

The authors did not compare a regression-based predictor using only conventional variables with one including the 70 gene panel. However the authors demonstrated the prognostic value of the 70 gene index using survival curves stratified by the NIH and St. Gallen criteria, which showed substantial separation between 70-gene prognostic groups that were either low or high risk by those conventional indices. These stratified survival curves also showed that optimal prediction was achieved when the gene index and conventional predictors were combined (Table 16, Appendix I, Evidence Tables 6, 7, and 9).

Buyse et al., 2006.⁵⁹ This study compared the MammaPrint assay with the conventional combination risk predictors Adjuvant Online, Nottingham Prognostic Index, and St. Gallen. Patients were drawn from five distinct European institutions, in the context of an independent, multicenter validation study performed by the TRANS-BIG consortium. Gene expression in frozen tumor specimens from node negative patients younger than 60 years old who did not

receive systemic adjuvant chemotherapy, and were diagnosed between 1980 and 1998 was characterized using the MammaPrint® assay. Final results were obtained for 302 out of 402 eligible patients. The median followup was 13.6 years, and the overall rate of distant metastasis was 25 percent.

The three primary end points of the study were time to distant metastases (TTM), DFS, and OS. The hazard ratios of the MammaPrint assay for TTM and OS were statistically significant after adjustment for St. Gallen, NPI and Adjuvant! On-line, but were generally far below (in the 1.5-2.5 range) that seen in the original validation cohort.^{25,58} The partial explanation offered by the authors was that this study had a longer median followup time than the one used by the van de Vijver²⁵ cohort. Additionally, the authors introduced an interesting analysis showing the marked (3-6 fold) lowering of the hazard ratio for various endpoints when patients were artificially censored at increasing times, up to 10 years. Also, none of the ER positive patients reported in this study received hormonal therapy as did some of the original van de Vijver²⁵ cohort.

Specificity and sensitivity of the MammaPrint assay and the Adjuvant! algorithm were compared for distant metastases within 5 years and for death within 10 years. Similar sensitivities were found, but a higher specificity was demonstrated for MammaPrint. The areas under the Receiver operating characteristic (ROC) curves were comparable between MammaPrint and Adjuvant! (0.68 vs. 0.66 for distant metastases at 5 years). The use of alternative thresholds for the Adjuvant! Online results did not change the overall results, and Adjuvant! hazard ratios were greater than unity but not statistically significant when adjusted for the gene signature. Finally, there was no statistical heterogeneity in any outcomes between centers, suggesting that this prediction model has transportability across populations with possibly different genotypic patterns.

This study is particularly important in that it provided the first evidence for the degree of clinical validity of the MammaPrint assay distinct from the 70-gene signature. It provided insight into the impact of differing lengths of followup in validation cohorts, and concluded that the prognostic contribution was sizable. However, this study's predictions were made in the context of no adjuvant hormonal or chemotherapy treatment, thus its applicability to women over 60 years old and treated with tamoxifen is unknown⁶⁸ (Table 16, Appendix I Evidence Tables 6, 7, and 9).

Glas et al., 2006.⁵⁸ This study used the same patients as in the van't Veer,²¹ and van de Vijver²⁵ studies and compared the currently offered MammaPrint assay results to the results of the previous studies. RNA was available for all the 78 patients in the van't Veer series, but only 145 lymph node negative patients were available for reanalysis from the van de Vijver series. A different reference RNA was used, as well as a different quantification method, however odds ratios and hazard ratios were very similar. A total of 15 patients were incorrectly classified into discrepant risk categories. The results of the 70-gene signature used in the original cohorts therefore apply equally to the MammaPrint assay based on that signature (Table 16, Appendix I Evidence Tables 6, 7 and 9).

H/I Ratio

A synopsis of the clinical validity evidence presented in the following section is reported in Table 18.

Ma et al., 2004.⁶⁴ This study reported the development of the two-gene ratio predictor. The authors generated gene expression profiles with gene chips from whole and laser-capture microdissected (LCM) frozen tumor specimens from 60 ER positive, node positive or negative breast cancer patients all treated with adjuvant tamoxifen monotherapy. Twenty-eight of the cohort (46 percent) experienced a distant recurrence within 4 years and 54 percent had no recurrence by 10 years. Twenty-two thousand genes were screened in the whole tissue sections and in LMC samples for their ability to predict DFS. Only three genes were highly predictive of DFS in both tissue sets, with over-expression of HOXB13 predicting recurrence and over-expression of IL17BR predicting non-recurrence. These expression values were combined in the form of a ratio, which outperformed both existing biomarkers and either gene alone. The univariate OR (interquartile) was 10.2 (95 percent CI, 2.9-36), multivariate OR was 7.3 (95 percent CI, 2.1-26.3) with adjustment for tumor size, PR and ERBB2 (none statistically significant) in a logistic regression. Area under the receiver-operating-characteristic curve (AUCs) for the ratio were reported in the 0.8 range.

Next, the above analysis was repeated using just the two-gene ratio calculated by RT-PCR on 59 fresh-frozen samples from the training set along with 20 additional FFPE specimens to independently validate the ratio. Sixteen of these 20 were accurately predicted. The RT-PCR-measured expression was reported to have similar predictive power to that measured via gene arrays. No comparison with the full array of clinical predictors (e.g. tumor grade) or with standard combination predictors (e.g., Adjuvant!) was performed (Table 18, Appendix I, Evidence Tables 10, 11, and 13).

Reid et al., 2005.⁶⁹ In this paper the authors attempted to validate the two-gene ratio on an independent cohort of 58 patients with ER positive breast cancer. These patients had been treated with tamoxifen monotherapy, had larger tumors, a higher frequency of lymph node metastases (78 percent vs. 47 percent), and a higher HER-2 positivity (21 percent vs. 5 percent) than those in the Ma et al., 2004 study. Eighteen patients had distant recurrences within a median time of 31 months, and 40 had no recurrence after a median of 93 months (range 70-125). The expression of the genes HOXB13 and IL17BR was measured by RT-PCR and the association between their expression and outcome was assessed by use of univariate logistic regression, AUC, a two-sample t test, and a Mann–Whitney test. None of these analyses revealed any statistical relationship with outcome.

The authors then took the original data of Ma et al.⁶⁴ and applied standard supervised methods to this and to another independent data set with 99 similar patients.⁷⁰ They tried to estimate the classification accuracy obtainable by using two or more genes in a microarray-based predictive model, using linear discriminant analysis and extensive cross-validation. The authors failed to validate the two-gene ratio and found high error rates with two-gene predictors.

Overall, findings from this paper argued against the prognostic utility of the two-gene ratio in ER positive breast cancer patients treated with tamoxifen. However, it must be noted that a different part of the transcripts were assayed in the two studies and that differences could be due to the documented differences in the populations used, which were neither clinically nor therapeutically homogeneous, with small validation sets⁷¹ (Table 18, Appendix I, Evidence Tables 10, 11 and 13).

Goetz et al., 2006.⁶² To investigate the prognostic performance of the two-gene ratio, this study analyzed FFPE samples from 206 ER-positive patients treated in the tamoxifen-only arm of a Phase III randomized trial of tamoxifen alone versus tamoxifen plus fluoximesterone conducted through the NCCTG (North Central Cancer Treatment Group).⁶⁴ RT-PCR expression

values for each gene were normalized using a standard curve (Appendix D) obtained by analyzing the human universal total RNA (Stratagen, La Jolla, CA), rather than the standard reference gene method, although the authors stated that control genes were not necessary to assess the expression ratio. The following end points were considered: RFS (time from randomization to any event of recurrence, contralateral breast cancer or death), DFS (time from randomization to any event of recurrence, or contralateral breast cancer, or other cancer, or death), and OS (time from randomization to death).

Cutoffs points that best predicted RFS, DFS and OS were identified: the optimal cut-off for the entire cohort was -1.85, corresponding to the 58th percentile, whereas the 59th percentile (-1.34) was used for the node-negative group (n = 130), and the 90th percentile (4.4) best discriminated in the node positive group (n = 86).

The ratio showed modest outcome prediction value in the entire cohort, with cross-validated hazard ratios near 1.5 and *P* values around 0.05, with the predictive value being restricted to the node-negative subset of patients (hazard ratios 1.7 to 2, *P* values = 0.04-0.06). In the node-positive group the ratio had no relationship to relapse or survival. The authors concluded that a high 2-gene expression ratio is associated with increased relapse and death in patients with node-negative, ER positive breast cancer treated with tamoxifen.

Overall this study provided some support of the two-gene ratio signature's prognostic value in ER positive, lymph node negative patients, but both the magnitude of that effect and the statistical support were modest, and the relevant cutoffs used for discrimination between high and low risk were optimized for each endpoint and patient subgroup. Hence, this is closer to a training than validation exercise (Table 18, Appendix I, Evidence Tables 10, 11 and 13).

Ma et al., 2006.⁶¹ This study examined a consecutive series of patients from Baylor University diagnosed between 1973 and 1993 with stage I or II breast cancer. The patients did not have distant spread, and non-relapsed cases had a median followup of 6.8 years. The authors reported data on the clinical validity of the two-gene expression index (HOXB13:IL17BR), which is the base of the H/I assay. A different normalization strategy (Table 1) from Ma et al., 2004⁶⁴ was applied to obtain this index. FFPE samples only yielded 852 analyzable cases out of 1,002 patients.

This population had 72 percent node negative, 73 percent ER positive, and 16 percent HER-2 positive patients, with an overall recurrence rate of 31 percent. A higher HOXB13:IL17BR index was associated with a higher risk of relapse (hazard ratio=1.5, *P*<0.001). In a stratified analysis, univariate Cox regression indicated that the HOXB13:IL17BR index was only significant in node-negative patients (hazard ratio = 1.6, *P*<0.001 vs. hazard ratio=1.2, *P*=0.1,) and further subsetting indicated that the interaction with node status was statistically significant for the HOXB13:IL17BR index (*P*= 0.02) only in ER positive patients. The HOXB13:IL17BR index correlated significantly with predictors of poor prognosis (i.e., HER-2 amplification, S-phase fraction, and number of positive lymph nodes) and correlated inversely with ER and PR expression.

The authors identified the optimal cut-off point for the index by analyzing a training set of ER-positive untreated patients (n=205), in order to obtain the smallest *P* value from a log-rank test in Kaplan-Meier survival analysis. The selected threshold (of about 1.0) was validated in a separate test set of untreated patients (n=103), and was also applied in the analysis of the tamoxifen-treated group of patients (n=122). Kaplan-Meier curves and univariate Cox regression analysis indicated that this cut point stratified patients into significantly different risk groups. Results from the Kaplan-Meier plots suggested that the prognostic power of the two-gene index

was independent of tamoxifen therapy. The hazard ratio obtained in multivariate Cox proportional hazards regression, incorporating age, tumor size, S-phase fraction, PR status, and tamoxifen therapy, confirmed the prognostic role of the HOXB13:IL17BR index (hazard ratio=3.9, 95 percent CI = 1.5 to 10.3, P value = 0.007), in ER positive, node negative, patients irrespective of tamoxifen treatment. The index was also demonstrated to be a continuous predictor of DFS in untreated patients. The authors concluded that the two-gene index was a significant predictor of clinical outcome in ER positive, node-negative, patients regardless of tamoxifen therapy.

This study validated the two-gene ratio gene expression profile, developed the two-gene index, and provided preliminary evidence for its prognostic value. Classification probabilities were not presented, and its incremental value over conventional predictors was not reported, although some components of such predictors were included in the multivariate analyses (Table 18, Appendix I, Evidence Tables 10, 11 and 13).

Jansen et al., 2007.⁷² This clinical study evaluated the ability of the HOXB13-to-IL17BR expression ratio to predict DFS in breast cancer patients treated with tamoxifen. The HOXB13 and IL17BR expression levels were measured by RT-PCR in 1,252 primary breast tumor patients and normalized with respect to 3 housekeeping genes⁷³. The study population was a mix of ER-positive (73 percent), lymph node-positive (52 percent), tamoxifen-treated (14 percent), and chemotherapy-treated (17 percent) patients, with additional patients treated with tamoxifen or chemotherapy after relapse (55 percent). Patients with ER-positive tumors with node negative primary breast cancer ($N = 468$) were followed for DFS. Patients with recurrent breast cancer treated with first-line tamoxifen monotherapy ($N = 193$) were followed for progression free survival (PFS). This study used different populations, protocols, normalization strategy, and ratio thresholds than Ma et al. 2006.⁶¹

The study evaluated the relation between the HOXB13-to-IL17BR ratio and tumor aggressiveness in lymph node negative, ER positive patients who did not receive adjuvant systemic chemotherapy ($N=468$). Of these patients, 46 percent had a relapse during the followup period. The HOXB13-to-IL17BR ratio, as a univariate continuous variable, was significantly associated with a poor DFS (hazard ratio=1.6, $P=0.02$) and a poor OS ($P<0.001$, data not reported). When traditional factors were added to the model, the HOXB13-to-IL17BR ratio continued to contribute significantly to DFS and OS prediction, either as a continuous variable or after dichotomization according to published pre-specified thresholds⁶¹ (Table 18).

The same analysis was performed on ER-positive, lymph node-positive tumors from untreated patients, who were mainly enrolled in the early 1980's ($n=151$). Univariate analysis of the continuous HOXB13-to-IL17BR ratio was associated with a poor DFS and a poor OS. In the multivariate model for this population, the index was significantly associated with OS (P value = 0.001), but less strongly with DFS (P value = 0.065). The dichotomized index was not related to DFS (data not shown).

Finally, the authors evaluated the prognostic performance of the HOXB13-to-IL17BR ratio in 193 ER-positive primary breast tumors in relapsed patients treated with first-line tamoxifen monotherapy. Both univariate and multivariate analyses revealed that the ratio, continuous and dichotomized, was strongly associated with PFS (Table 18).

This study is by far the largest done so far concerning the potential value of the 2-gene ratio. It provided evidence of the clinical validity of the HOXB13-to-IL17BR ratio in ER positive, node negative patients who did not receive systemic adjuvant therapy, and also in ER positive relapsing patients whose relapse was treated with tamoxifen. However, the study was calculated

and dichotomized in a somewhat different manner than in Ma et al., 2006.⁶¹ Additionally, comparisons were not provided with conventional combination risk indices, nor were classification probabilities provided for the models with and without the ratio. Therefore, incremental predictive values could not be accurately assessed. Although qualitative conclusions are not affected, there are some differences between the quantitative results reported in the text and tables (Table 18, Appendix I, Evidence Tables 10, 11 and 13).

Jerevall et al., 2007.⁶³ In this paper the authors investigated whether the two-gene ratio can predict the benefit of 2 versus 5 years of tamoxifen treatment in postmenopausal breast cancer patients, and also predict the ratio's prognostic value in systematically untreated pre-menopausal patients. Expression of HOXB13 and IL17BR were quantified by RT-PCR in tumors from 264 randomized postmenopausal patients and 93 systemically untreated premenopausal patients. The two study populations were collected as part of a collaborative study between two centers in Sweden, and 72 percent of the randomized patients were lymph node positive and 74 percent ER positive. To stratify the patients into risk groups the authors dichotomized the ratio using the median, a procedure and dichotomization differing from the approach used by Ma 2006.⁶¹ The results from the prediction of prolonged treatment benefit are reported under Key Question 4, Clinical Utility.

The ratio proved to be significantly correlated to tumor size, ER, PR, HER-2, Nottingham histologic grade (NHG), ploidy, and S-phase. ER, HER-2, S-phase and NHG correlations were mostly due to IL17BR, while PR and ploidy correlations showed contribution from both genes. The authors concluded that a lower expression of IL17BR, but not HOXB13, was correlated to several factors related to poor prognosis, and thus IL17BR might be an independent prognostic factor in breast cancer, and that HOXB13 may be correlated with tamoxifen resistance. However, the ratio had no prognostic value in ER negative postmenopausal patients and they were excluded from subsequent analyses.

In summary, this study produced additional developmental evidence of the prognostic value of the HOXB13-to-IL17BR ratio, and of the two individual genes, in ER positive breast cancer patients who received systemic adjuvant therapy. However, neither the patient profile nor the mode of calculation of the ratio were identical to previous studies, and the results differed from previous reports, as the ratio predicted for worse outcome in lymph node positive patients (Table 18, Appendix I, Evidence Tables 10, 11 and 13).

Key Question 4. What is the clinical utility of these tests?

The clinical utility of a test tells us whether the test helps discriminate between those who will have more or less benefit from a therapeutic intervention. This can only be assessed in the context of randomized clinical trials, where benefit can be measured in terms of an improvement of clinical outcomes such as overall survival, disease-free survival, chemotherapy toxicity, or quality of life.

The prognostic estimates provided in the previous section, however—have a relationship to clinical utility—providing an upper limit on the degree of clinical benefit that can be provided by chemotherapy for a given endpoint. For example, if the 10-year cancer recurrence rate without adjuvant chemotherapy is estimated to be 5 percent, the maximum absolute benefit to be derived from chemotherapy cannot exceed 5 percent. Furthermore, knowledge that chemotherapy generally only prevents a minority of recurrences tells us that the absolute benefit in terms of recurrence in that situation will be likely less than 2 percent. So while prognostic estimates are

not direct estimates of benefit per se, they provide enough information that could be used to crudely estimate benefit and be sometimes relevant for patient decision-making.

Oncotype DX

Currently a prospective randomized clinical trial, TAILORx, is underway with the goal of assessing the value of adjuvant chemotherapy among patients with mid-range RS results. However, one other published study does address the potential value of the RS in predicting chemotherapy benefit.

A synopsis of the clinical utility evidence presented in the following section is reported in Table 19.

Paik et al., 2006.⁵³ The authors used the Oncotype DX assay to investigate whether the RS was a predictor of the benefit from chemotherapy in ER-positive, lymph node negative, breast cancer patients. This study used 651 patients from the NSABP B-20 randomized trial and compared a group treated with both tamoxifen and chemotherapy with a group of patients who were randomized to tamoxifen only. Gene expression analysis was found to be correlated with chemotherapy benefit, defined in terms of 10-year distant recurrence-free survival (DRFS).

Kaplan-Meier analysis on all patients showed a significant benefit from the use of chemotherapy (P value = 0.02), however when the data was stratified by RS risk groups, only the high RS risk group of patients benefited from using chemotherapy (P value = 0.001).

When the authors used multivariate Cox proportional hazard analysis, findings about the benefit from chemotherapy use were unclear due to large confidence intervals in the low and intermediate RS risk groups (low RS risk group, RR=1.31; 95 percent CI: 0.46–3.78; intermediate RS risk group, RR = 0.61; 95 percent CI, 0.24 to 1.59). Patients classified in the high RS risk group, however, showed a significant benefit from the use of chemotherapy (RR=0.26; 95 percent CI: 0.13–0.53).

The authors also looked for interaction between each variable and chemotherapy treatment using separate likelihood ratio tests. The RS was the only significant interaction ($P=0.038$), with only slight statistical weakening when age, tumor size, tumor grade and site were added to the model individually (P values from 0.035 to 0.068). When RS was fit as a continuous score, there was not a clear threshold that predicted no benefit for chemotherapy.⁵³

Overall, this study produced preliminary, high-quality evidence that the RS from the Oncotype DX assay has clinical utility, i.e. predictive power in assessing the benefit of chemotherapy usage in ER-positive, lymph node negative breast cancer patients. The embedding of this study within a large, well conducted RCT was a strength. However, some patients from the tamoxifen-only arm of the NSABP B-20 trial were in the training data sets for the Oncotype DX assay. While the algorithm was trained for the outcome of recurrence and not chemotherapy benefit, optimization of recurrence prediction in one arm of this study could translate into a somewhat enhanced estimate of chemotherapy benefit, although it is unlikely to account for the large effect seen here. Finally, while the models could not sustain the inclusion of all possible clinical variables, they could have included a composite score, either standard risk predictors, or one tailored for the data set (Table 19, Appendix I, Evidence Tables 1, 2 and 4).

Correlation between RS and chemotherapy response.

*Gianni et al., 2005.*⁴⁹ This study focused on the complete pathological response (pCR) to preoperative chemotherapy in node negative and positive patients, looking at the correlation between pCR and RS. Two independent cohorts of patients were used, the cohort from the Italian

National Cancer Institute of Milan, Italy, and the M.D. Anderson Cancer Center cohort from the M.D. Anderson Cancer Center of Houston, U.S. (Appendix I, Evidence Table 2), and were evaluated by two different technologies (RT-PCR and the Affymetrix hgu133a array). The study also identified additional genes that are associated with pCR and allowed the development of a new gene panel associated with pCR, as well as the evaluation of the association of *Oncotype DX RS* with pCR.

Results of the Oncotype DX assay in the Milan cohort. Three hundred and eighty-four genes were analyzed by RT-PCR in the Milan cohort of patients, including the 21 genes assessed by the *Oncotype DX* assay. Data showed good discrimination of pCR by RS. Probit regression-based models with and without the incorporation of the RS resulted in a *P* value of 0.005 in a global likelihood ratio test.

Preliminary evidence that the RS from the *Oncotype DX* assay has predictive power in assessing the likelihood of pCR after pre-operative chemotherapy was obtained in this study. (Table 19, Appendix I, Evidence Tables 1, 2 and 4).

Mina et al., 2006.⁵¹ In this study paraffin-embedded pre-treatment core biopsies from a completed phase II trial of 70 patients with newly diagnosed stage II or III breast cancer who were treated with sequential doxorubicin and docetaxel were used to identify genes that correlate with response to pCR. Gene expression was investigated by RT-PCR in 45 patients, using the same procedures of the *Oncotype DX* assay. A total of 192 genes (187 candidate genes and 5 reference genes) were tested, including those used to compute the *Oncotype DX* Recurrence Score.

Individual genes, as well as groups of biologically related genes, were found to be associated with pCR, however no correlation between *Oncotype DX RS* and pCR was found (*P* = 0.67). A total of 22 individual genes had an uncorrected *P* value of less than 0.05 in a likelihood ratio test derived from logistic regression models; however 13 genes would be expected to correlate with pCR at the *P* value level of 0.05 level by chance alone.

This study provides preliminary evidence that the RS from the *Oncotype DX* assay cannot predict pCR after primary chemotherapy in advanced breast cancer patients (with variable ER and HER-2 status, lymph node involvement, tumor size, and tumor grade) (Table 19, Appendix I, Evidence Tables 1, 2 and 4).

Chang et al., 2007.⁵⁵ This study is currently in press for Breast Cancer Research Treatment. The authors investigated if expression of the 21 genes of the *Oncotype DX* assay and other candidate genes in locally advanced breast cancer tumors could be used to predict response to docetaxel treatment. The 97 women in this study were diagnosed and were enrolled into three phase II studies with the neoadjuvant docetaxel at Baylor College of Medicine, Houston, U.S. Clinical response was assessed by Response Evaluation Criteria in Solid Tumors (RECIST) criteria: clinical complete response (CR) was defined as complete disappearance of the tumor, while partial response (PR) was defined as at least 30 percent decrease in unidimensional size. An increase of more than 25 percent was defined as clinical progressive disease (PD). Any response that did not meet the definition of CR, PR, or PD was defined as stable disease (SD). All patients received primary surgery and standard adjuvant therapy. Core biopsies from 97 patients were obtained before treatment and RNA levels of expression for the selected genes were studied by RT-PCR, following the specified protocols for the *Oncotype DX* assay.

Of the selected 97 patients, 81 (84 percent) had sufficient invasive cancer, 80 (82 percent) had sufficient RNA to perform the RT-PCR based assay, and 72 (74 percent) had known clinical response data. The mean age was 48.5 years, while the median tumor size was 6 cm. A clinical

CR was observed in 12 patients (16.7 percent) a partial response in 41 (56.9 percent), a stable disease in 17 (23.6 percent), while progressive disease was present in 2 patients (2.8 percent). Pathologically, pCR was observed in 2 patients (3.2 percent), ‘incomplete’ responses were observed in 61 patients (96.8 percent), and pathologic response was unknown for 9 patients.

The authors found that a CR was more likely associated with a high RS ($P = 0.008$). When the RS was used as continuous variable, a 50 unit increase in the RS was associated with a five-fold increase in the odds of achieving clinical CR (95 percent CI 1.3, 6.0). Moreover, the logistic model for the RS indicated that a 14-unit increase in the RS (the difference between low and high risk groups, as defined by the standard thresholds) was associated with a complete clinical response odds of 1.7 (95 percent CI 1.15, 2.60). The authors concluded that a high risk patient is at least 1.7 times more likely to achieve a clinical CR with neoadjuvant chemotherapy compared to a low risk patient. Finally, the accuracy of the *Oncotype* DX RS in predicting the response to adjuvant chemotherapy with docetaxel throughout the range of RS values was judged to be at least moderate, with AUC of 0.73.

Overall, this study provided preliminary evidence that the RS from the *Oncotype* DX assay has predictive value in assessing the likelihood of a clinical CR to primary chemotherapy with docetaxel. However the small cohort patients points to the need for further confirmation (Table 19, Appendix I, Evidence Tables 1, 2 and 4).

***Oncotype* influence on decisionmaking.**

*Oratz et al., 2007 (in press).*⁵⁶ This study investigated whether the *Oncotype* DX RS had influenced both clinicians’ treatment recommendations and the actual treatment administered in patients with ER positive, lymph node negative, early (stage I or II) breast cancer. A retrospective analysis was performed on 74 patients from a community-based oncology practice for whom RS was determined. Treatment recommendations prior to RS knowledge were compared with treatment recommendations after RS knowledge, and to the treatment eventually administered.

Knowledge of RS changed the clinicians’ treatment recommendations in 21 percent of patients, and the actual administered treatment in 25 percent of the patients. In particular, the decision to add chemotherapy to the hormonal therapy was generally associated with the high-risk group, whereas the decision to change from chemotherapy to hormonal therapy was associated, in general, with low RS.

While this study produced preliminary evidence that knowledge of the RS from the *Oncotype* DX assay can have an impact on the clinical management of patients diagnosed with ER positive, lymph node negative, early breast cancer, it did not report specifically what the patients (or doctors) were told or understood about their risk of recurrence. Because it is unknown whether absolute risks were a factor in decision-making, the study is minimally informative as to the actual risk thresholds used by women and their treating physicians (Appendix I, Evidence Tables 1, 2 and 4).

Economic studies.

*Hornberger et al., 2005.*⁶⁷ The objectives of this study were twofold. First, the authors sought to estimate the incremental benefits, costs, and cost-effectiveness of using *Oncotype* DX to better assign risk of distant recurrence-free survival associated with early stage breast cancer. Secondly, the authors wanted to assess the factors that most influence potential benefits and efficient use of the 21-gene RT-PCR recurrence score. The outcomes of interest to the study included overall survival, relevant costs of breast cancer care, and distant recurrence-free survival.

Cost-utility analyses used a Markov model to forecast overall survival, quality of life, costs, and cost-effectiveness. Two scenarios were considered, based on NCCN classification of patients with lymph node negative, ER positive, early stage breast cancer who were expected to receive 5 years of hormonal therapy into a low risk (T1a N0-1mi) group that did not receive chemotherapy versus a high risk (T1b with unfavorable features or T1c) group that did receive chemotherapy. Patients were then reclassified using the RS. Annual risks of recurrence and survival were obtained from published meta-analyses of clinical trials, and the study model included costs of the assay and drugs, including chemotherapy (Table 20, Appendix I, Evidence Tables 1, 2, and 5).

Summary of study findings. The analysis reported that using the 21-gene RT-PCR assay to reclassify patients who were defined by NCCN criteria as low risk (to intermediate or high risk) would lead to an average gain in overall survival per reclassified patient of 1.86 years. Total cost estimates increased by about \$25,000. This amount included \$12,190 to identify intermediate- or high-risk patients and at least \$15,000 for chemotherapy, and was offset by savings of \$2,344 because of the lower risk of recurrence. The cost-utility of RS testing for this cohort was \$31,452 per quality-adjusted life-year (QALY) gained.

The authors also reported that reclassifying patients defined as high risk (by 2005 NCCN criteria) to low risk (using the 21-gene RT-PCR assay) was cost saving. The added cost of testing (\$7,073) to identify 1 reclassified patient was offset by an estimated \$15,000 in savings for eliminating the need for chemotherapy.

Using the 21-gene RT-PCR assay was expected to improve quality-adjusted survival by a mean of 8.6 QALYs and reduce overall costs by about \$203,000 in a hypothetical population of 100 patients with characteristics similar to those of the NSABP B-14) participants, more than 90 percent of whom were NCCN-defined as high risk. The estimated cost-effectiveness was most influenced by the propensity to administer chemotherapy based on the RS, and by the very small proportion of patients at low risk as defined by 2005 NCCN guidelines. The 2007 NCCN guideline indicates that the use of chemotherapy in these patients is now considered optional, thereby diminishing the utility of this model.

Critical appraisal of the analysis. The EPC team appraised the analysis using published guidelines for good practice in decision-analytic modeling in health technology assessment, Philips 2004.⁴¹ The appraisal took into consideration the domains of structure, data, and consistency (Table 20, Appendix I, Evidence Table 5).

Structure and Data. The authors provided a clear description of many aspects of the structure of the analysis, including the decision problem, objectives of the evaluation, perspective of the analysis, rationale for the model structure, and structural assumptions. However, the model inputs were not entirely consistent with the stated perspective of the analysis. For instance, the model did not include all costs that are relevant from a societal perspective such as decreased productivity and days lost from work. Also the authors did not address the limitations in how utility estimates were derived. This is an important limitation because utility estimates can vary a lot depending on the methods that are used to derive the estimates. The authors also did not justify extrapolating beyond the 10-year followup period for which recurrence data is available. Finally, the authors did not report much information about their assessment of methodological and structural uncertainties. Without such information it is difficult to determine how their projections might differ if different assumptions were made in the decision model.

The authors correctly pointed out that the 2005 version of the NCCN breast cancer guideline recommends chemotherapy for all node-negative tumors greater than 1 cm (T1a).⁷⁴ Since 84

percent of the patients included in the Paik study²⁸ had tumors larger than 1 cm (T1c), it is unsurprising that a very large proportion of patients overall would be spared chemotherapy (gene expression profiling data expected to identify approximately half of these patients to have a low RS). However, by 2007 the NCCN panel had refined its criteria for recommending chemotherapy⁶, now considered optional (adjuvant hormonal therapy ± chemotherapy) for those with ER-positive HER-2-negative disease and tumors greater than 1cm (T1c). Therefore, it is reasonable to speculate that approximately half of these patients might opt for no chemotherapy. This is a similar proportion of patients that would be found to have a low RS, although these two groups of patients may not necessarily be the same.

Consistency. Appendix I, Evidence Table 5 notes that the authors did not report information about the internal and external consistency of their analysis. The analysis would be more convincing if it gave more information on whether the mathematical logic of the model had been tested (internal consistency) or if results from other models were available for comparison (external consistency). Nevertheless, the results of the model make intuitive sense and seem to be consistent with published data on the performance characteristics of the 21-gene RT-PCR recurrence score.

Summary of critical appraisal. Overall, the EPC team concluded that this economic analysis met most of the standards set by the rigorous guidelines of Phillips et al., 2004⁴¹. It is not clear whether the limitations noted above biased the results for or against the 21-gene RT-PCR assay, but extension of the timeframe beyond 10 years could overstate the benefits of using the assay. Given that this study was sponsored in part by the manufacturer of the 21-gene RT-PCR assay (Genomic Health, Inc., Redwood City, California), the EPC team would have had more confidence in the results if the authors had provided more information about methodological and structural uncertainties as well as other potential sources of bias such as the derivation of the utility estimates. The generalizability of these results to patients in 2007 is also limited, as the 2005 NCCN guidelines have since been updated. Thus, the team has only moderate confidence that the results of the economic analysis provide reasonable estimates of the potential cost-effectiveness of using the 21-gene RT-PCR assay to guide treatment of early stage breast cancer

Lyman et al., 2007.⁷⁵ The main objective of the second study⁷ was to estimate the cost-effectiveness of 21-gene RT-PCR assay-guided treatment of patients with ER positive, lymph node-negative, early-stage breast cancer with either tamoxifen alone or the combination of chemotherapy and tamoxifen.

This analysis incorporated data that validated the prognostic accuracy for distant RFS using a 21-gene RT-PCR assay in 668 lymph node-negative, ER positive women with early-stage breast cancer receiving tamoxifen on NSABP B-14. The analysis also incorporated data that validated the predictive accuracy for treatment efficacy in 651 patients randomized in NSABP B-20, and 645 patients in NSABP B-14.

The study design involved cost-utility analyses using a “clinical decision model” designed to compare clinical, economic, and quality of life outcomes for three adjuvant treatment strategies: 1) tamoxifen alone, 2) chemotherapy followed by tamoxifen, or 3) therapy based on the results of the 21-gene RT-PCR assay. Using the RS from the 21-gene RT-PCR assay, patients were classified as high risk ($RS \geq 31$), intermediate risk ($RS 18-30$), or low risk ($RS < 18$) for distant recurrence at 10 years. The third strategy assumed that low-risk patients would receive tamoxifen, and intermediate or high-risk patients would receive chemotherapy and tamoxifen. Clinical outcomes were estimated in terms of life expectancy or life-years saved as derived from NSABP B-20 and B-14 data. Economic outcomes included selected costs of cancer care,

including the costs of chemotherapy, surveillance without recurrence, use of the 21-gene RT-PCR assay, and treatment of recurrence. Quality of life outcomes were estimated based on the utility associated with use of chemotherapy. The treatment strategies were compared in terms of the additional cost of one strategy over another (marginal cost), the additional clinical benefit (marginal efficacy), and the additional quality-adjusted clinical benefit (marginal utility) (Table 20, Appendix I, Evidence Tables 1, 2 and 5).

Summary of study findings. The lowest expected mean cost per life-year saved was associated with treatment with tamoxifen alone (\$11,890), whereas the greatest expected mean cost was associated with treatment with both chemotherapy and tamoxifen (\$18,418). The expected cost of each strategy increased as the assumed cost of treating distant recurrence increased. Above a cost of \$100,759 for treating recurrence, therapy guided by the RS provided a net cost savings compared with other strategies and was always cost-saving compared with the chemotherapy and tamoxifen strategy. The tamoxifen strategy was associated with the lowest costs for all reasonable followup cost assumptions among those without recurrence. Therapy guided by the RS was favored over chemotherapy and tamoxifen for total chemotherapy costs exceeding \$5,822. The use of therapy guided by the RS was more costly for low-cost chemotherapy regimens not requiring additional supportive care, whereas a net cost savings between \$500 and \$10,000 was estimated with RS guided therapy for other commonly used and higher-cost adjuvant chemotherapy regimens.

Compared to tamoxifen alone, the expected incremental cost associated with RS-guided therapy was \$4,272. The expected incremental cost associated with chemotherapy and tamoxifen was \$6,527. The incremental cost-effectiveness ratio compared with tamoxifen alone favored the use of RS-guided therapy (\$1,944 per life-year saved) over the use of chemotherapy and tamoxifen (\$3,385 per life-year saved). When the analysis considered increases in healthy life expectancy, the incremental life-years saved increased for the RS-guided therapy compared with tamoxifen alone, and the corresponding marginal cost-effectiveness decreased.

Expected QALYs favored RS-guided therapy over chemotherapy and tamoxifen for all health utility values, with increasing incremental QALYs as the impact of chemotherapy on measured utility increased. Recurrence-score-guided therapy had greater expected QALYs compared with tamoxifen alone, until the utility associated with chemotherapy fell below 0.80. At a utility of 0.90 for adjuvant chemotherapy, RS-guided therapy was associated with a gain of 0.97 QALYs, a cost-utility ratio of \$4,432 per QALY compared with tamoxifen alone, and a gain of 1.71 QALYs with net cost savings when compared with the chemotherapy and tamoxifen combination.

Critical appraisal of the analysis. The EPC team appraised the analysis using published guidelines for good practice in decision-analytic modeling in health technology assessment Phillips 2004⁴¹, taking into consideration the domains of structure, data, and consistency (Table 20, Appendix I, Evidence Table 5).

Structure. Although the authors provided a clear description of the decision problem, they did not state the perspective of the model. Moreover, the authors did not provide enough information about the structure of the model to allow an evaluation of the appropriateness of the model type or of the causal relationships described by the model. The authors also did not justify extrapolating beyond the 10-year period for which recurrence data is available.

Data. The authors provided some explanation and justification of the data used in the analysis, citing previous work for some of the details. However, the authors did not include all relevant costs. They included the costs of adjuvant chemotherapy, surveillance, use of the

Oncotype DX assay, and treatment of recurrence, but they did not include other treatment-related direct costs (e.g., costs of administration, associated testing, and transportation) or indirect costs (e.g., decreased productivity). Although indirect costs may be implicitly included in utility values assigned to relevant health states, the authors did not provide enough information to determine whether that was done. The analysis would have been stronger if it had estimated cost-effectiveness with and without inclusion of indirect costs and other treatment-related costs. The authors did not mention any health-state utilities other than the utility with chemotherapy, and did not give sufficient detail about how they estimated the utility with chemotherapy. In addition, the authors did not report on the quality of the data. A single study was used as the source of estimates for the relative effects of the treatment strategies. The authors also did not report sufficient information about the sensitivity analysis and alternative assumptions. Finally, the authors did not report much information about their assessment of methodological and structural uncertainties.

Consistency. The authors did not report information about the internal and external consistency of their analysis, but the results of the model make intuitive sense. Generally, the results seem to be consistent with the cited data on the performance characteristics of the 21-gene RT-PCR RS.

Summary of critical appraisal. Overall, the EPC team concluded that this economic analysis did not meet many of the standards set by the rigorous guidelines of Phillips et al., 2004⁴¹. These limitations are particularly serious because the authors received research support from the manufacturer of the 21-gene RT-PCR assay. Consequently, the EPC team has little confidence in the results of this analysis.

Summary of available studies. Based on the evidence from the stronger of the two available studies, the EPC team concluded that the 21-gene assay, when used to guide treatment for patients previously classified as low risk by NCCN-defined criteria, may be cost-effective compared to standard treatment approaches in women with lymph node-negative, ER positive early-stage breast cancer. Similarly, the EPC team concluded that the 21-gene assay, when used to guide treatment for patients previously classified as high risk by NCCN criteria, may be cost-saving compared to standard treatment. The overall body of evidence on economic outcomes is weak because of the limitations of the two available studies.

MammaPrint

No published studies evaluated the ability of the 70-gene signature for the main MammaPrint assay to predict chemotherapy benefit.

Economic studies.

*Oestericher et al., 2005.*⁷⁶ The main objective of this study was to compare the cost-effectiveness of the Netherlands Cancer Institute gene expression profiling (GEP) assay to the NIH guidelines for the identification of early stage breast cancer patients who would benefit from adjuvant chemotherapy based on risk of distal recurrence. Although the references cited for the performance characteristics of the GEP assay indicate that the investigators were using data on MammaPrint, the article does not clearly state that they were analyzing MammaPrint.

The study design involved a cost-utility analysis. Using a Markov model, the investigators estimated the incremental cost and QALYs associated with use of the GEP assay as compared to use of the NIH guidelines in a hypothetical cohort of premenopausal women averaging 44 years of age newly diagnosed with stage I/II breast cancer. The performance characteristics of the tests

were based on data from the Netherlands Cancer Institute cohort.²⁵ In the Markov model, the investigators assumed that the results of the GEP assay would be used to classify patients as having a “good prognosis” or a “poor prognosis” based on a test cutoff derived from the first validation study of the GEP assay.²¹ They also assumed that the NIH guidelines would be used to classify patients as having a “good prognosis” or a “poor prognosis,” that women with a “poor prognosis” would receive adjuvant chemotherapy, and that women with a “good prognosis” would not receive chemotherapy. The model considered the following clinical events: distant recurrence of breast cancer, mortality due to distant recurrence, and mortality from other causes. The economic outcomes included the cost of the GEP assay, the cost of adjuvant chemotherapy, and the cost of managing distant recurrence of breast cancer. Quality of life outcomes were estimated in terms of QALYs, with utility estimates for specific health states derived from previous publications. The two strategies were compared in terms of the number of cases of distant recurrence prevented, costs, and QALYs. (Table 20, Appendix I, Evidence Table 5).

Summary of study findings. The NIH guidelines identified 96 percent of the cohort as high risk whereas the GEP identifies 61 percent of patients as high risk with sensitivities of 98 percent for the NIH guidelines and 84 percent for GEP. Specificities were 51 percent for GEP and 5 percent for the NIH guidelines. Since there is a 35 percent risk reduction in distant recurrence from use of chemotherapy, using NIH guidelines to identify high-risk women and treat with chemotherapy prevented 34 percent of distant recurrences compared to 29 percent for GEP. After including the negative impact on life expectancy and quality of life from chemotherapy and distant recurrence, the NIH guidelines and GEP yielded 10.08 and 9.86 QALYs respectively. Total costs were \$32,636 for the NIH guidelines and \$29,754 for GEP.

Although the GEP assay was projected to identify 35 percent fewer women for chemotherapy than NIH guidelines, quality of life benefits in the women who did not need chemotherapy were outweighed by the decrease in life expectancy in the women who needed chemotherapy but did not receive it because of GEP’s lower sensitivity.

The authors concluded that, in order to improve quality of life by allowing women to safely avoid chemotherapy while not missing women whose survival is compromised by avoiding therapy, GEP’s sensitivity would have to increase to at least 95 percent while maintaining a specificity of 51 percent. The GEP assay did not attain a sensitivity of 95 percent regardless of the test cutoff used in the analysis.

Critical appraisal of the analysis. The EPC team appraised the analysis using published guidelines for good practice in decision-analytic modeling in health technology assessment,⁴¹ taking into consideration the domains of structure, data, and consistency (Table 20, Appendix I, Evidence Table 5).

Structure. As indicated in Table 20, the authors provided a clear description of most aspects of the structure of the analysis, including the decision problem, objectives of the evaluation, perspective of the analysis, rationale for the model structure, and structural assumptions. The model inputs were consistent with the stated perspective of the analysis. The authors did not justify using a timeframe beyond the 6.7-year period for which recurrence data is available.

Data. The article was very strong in providing explanation and justification of the data used in the analysis. Limitations were that the authors did not justify extrapolation of data beyond 6.7 years of followup and that they only compared their model to the NIH guideline. In addition, although the authors listed a number of references for their use of utilities, they did not provide any explanation of how they derived specific utility estimates from these references. They also did not provide any explanation of the methods or scaling techniques that were used to derive the

utility estimates. Thus, we can not determine whether the utilities were based on the standard gamble techniques, which is the gold standard, or on other scaling techniques. This is important because the standard gamble techniques generally yields utility values that are higher than the values derived using other techniques. The estimates used in this study seem low compared to the values assigned to most serious health conditions.^{77,78} Also, these references for the utility estimates are significantly more dated than some of the references used to obtain cost data.

Consistency. The authors discussed the internal and external consistency of their analysis, and the results of the model make intuitive sense.

Summary of critical appraisal. Overall, the EPC team concluded that this economic analysis met most of the rigorous standards set by Phillips et al., 2004.⁴¹ The EPC team therefore has confidence in the results of this analysis. Although we had some uncertainty about the utilities used in the analysis, the EPC team believes that this limitation is unlikely to have changed the overall conclusion of the authors, which is based on the lack of sensitivity of the GEP assay.

H/I Ratio

Jerevall et al., 2007.⁶³ This paper investigated whether the two-gene ratio can predict the benefit of 2 years versus 5 years of tamoxifen treatment in postmenopausal breast cancer patients, and also predict the prognostic value in systematically untreated premenopausal patients. Expression of HOXB13 and IL17BR were quantified by RT-PCR in tumors from 264 randomized postmenopausal patients and 93 systemically untreated premenopausal patients. The two study populations were collected as part of a collaborative study between two centers in Sweden, and 72 percent of the randomized patients were lymph node positive and 74 percent ER positive. To stratify the patients into risk groups the authors dichotomized the ratio using the median. Thus the normalization procedure and dichotomization differed from the approach used by Ma.⁶¹ The prognostic results from this study are reported under Key Question 3 (clinical validity).

Kaplan-Meier analysis of data from postmenopausal ER-positive patients demonstrated that a low HOXB13-to-IL17BR ratio was associated with a benefit to receiving 5 vs. 2 years of tamoxifen treatment (univariate $P=0.021$; in KM analysis). There was no benefit ($P=0.9$) in patients who had a high ratio, which mainly appeared due to the low expression of HOXB13 genes ($P=0.010$, in Kaplan-Meier analysis). The predictive significance of both the two-gene ratio and the HOXB13 gene alone was maintained using a Cox proportional hazard modeling, adjusting for tumor size, PR status, and lymph node status.

The authors concluded that the ratio, or even HOXB13 alone, could predict the benefit of prolonged endocrine therapy, and that a lower expression of IL17BR, given its correlation to poor prognosis, could be an independent prognostic factor.

Neither the patient profile nor the mode of calculation of the ratio were identical to previous studies (Table 21, Appendix I, Evidence Tables 10, 11 and 13). However this study produced additional developmental evidence about the prognostic utility of the HOXB13-to-IL17BR ratio, and of the two individual genes, in ER positive breast cancer patients who received systemic adjuvant therapy.

Ongoing Studies

TAILORx (Trial Assigning Individualized Options for Treatment (Rx)).

The primary objective of TAILORx is to compare the DFS of women with previously-resected axillary-node-negative breast cancer who have an *Oncotype* DX RS of between 11 and 25 when treated with both adjuvant chemotherapy and hormonal therapy versus hormonal therapy alone. It should be noted that this range is lower on both ends than the standard “Intermediate” RS range, viz. 18-30. This represents a more conservative approach to the use of the RS than is suggested by current categories, in that subjects who agree to forego chemotherapy in this trial will be at lower risk than those in the current “low risk” RS group. The secondary objective is to determine if adjuvant hormonal therapy alone is sufficient treatment (i.e., 10-year distant DFS of at least 95 percent) for patients with an RS of less than or equal to 10.

This study will not provide direct evidence for the value of *Oncotype* DX, as all patients in the trial will receive the test. The trial results will indicate whether adjuvant chemotherapy is of value within the trial’s intermediate RS range, and will serve as further validation of the absolute risk of recurrence in subjects with scores above and below the range. This will provide better estimates of the degree of benefit from utilization of the test, but will not directly examine what therapeutic choices would have been made and clinical outcomes incurred if only standard risk prediction tools were used. However, since standard risk prediction indices will be calculable, that information may be inferred. First results from this trial are expected in approximately 2013.

MINDACT (Microarray for Node-Negative Disease may Avoid Chemotherapy).

MINDACT is a multi-center, prospective, phase III randomized study comparing use of the MammaPrint assay with a common clinical-pathological prognostic tool, Adjuvant! Online, to select patients for adjuvant chemotherapy in node-negative breast cancer. Patients at low risk by both MammaPrint and standard clinical-pathological criteria will not receive chemotherapy, patients at high risk by both criteria receive chemotherapy, and patients with discordant criteria will be randomized to use either MammaPrint only or standard criteria to decide treatment (i.e., randomized to receive adjuvant chemotherapy or not). This will directly test whether the choice of chemotherapy guided by MammaPrint provides benefit over that guided by the Adjuvant! criteria.

Other Relevant Studies

Fan et al., 2006.⁷⁹ No key questions relevant to the evaluation of gene expression-based prognostic estimators was directly addressed in this study, but the agreement between gene-expression tests and other predictors was evaluated, as well as their individual performance on a common dataset. In particular, the 70-gene signature, the gene panel used in *Oncotype* DX, the 2-gene ratio, and other gene expression signatures were considered. This investigation was carried out on the 295 samples from stage I-II breast cancer patients, which had been used to

develop the 70-gene test²¹. The *Oncotype* DX RS and the 2-gene ratio were estimated from microarray gene expression data (i.e., not RT-PCR), and thus were not obtained according to the protocols and methods used in the corresponding marketed assays. These are therefore described as “derived” scores below.

All tests except the 2-gene ratio (hazard ratio of about 1) were highly significant predictors of OS and DFS. The agreement between MammaPrint and derived RS was 81 percent (239/295). However the intermediate and high risk groups, as defined by the RS gene panel, were considered as one group in this paper and compared to the poor prognosis group of patients, as defined by the MammaPrint signature. ER status, tumor grade, tumor size, and lymph node involvement also proved to be significant univariate predictors. The coefficients of clinical predictors were allowed to vary between models in this analysis. All the analyses were repeated for the ER positive (N=225) subset with qualitatively similar results. Good, but not perfect correlation between predictions was found. This was surprising since classification was obtained using different gene sets. The degree of prediction over and above “standard” clinical stratifiers was not clear in the paper and the reclassification of samples was not done.

This study is of interest since it compared 5 different classifiers. However, it should not be regarded as a validation of either the *Oncotype* DX or the H/I ratio assays, since actual tests were not used on these patients and the RS and the two-gene index estimates were obtained from microarray data. In addition, since this was the same dataset used in the development of the 70-gene signature, it would be expected to perform better than the RS, for which this was a true test set.

Espinosa et al., 2005.⁸⁰ In this paper the authors developed an RT-PCR based version of the 70-gene expression signature^{21,25} RT-PCR was used to measure, in breast cancer biopsy specimens, the expression of the 70-gene signature, as well as four additional genes (HER-2, EGFR, PLAT, and MUC-1) related to prognosis. The study population was 96 patients diagnosed between 1991 and 1997 for whom samples and followup were available and who were seen in a single Madrid hospital. Half of the patients were lymph node positive, 75 percent ER positive, and 25 relapses were observed after a median of 70 months of followup. Eighty percent of ER positive patients received tamoxifen, and 74 percent of patients overall received adjuvant chemotherapy.

The objective of the authors was to reproduce the results obtained with the 70-gene profile through an alternative technology. However, for technical reasons only 60 of the 70 genes could be investigated. For this reason, the study cannot be considered a validation of the 70-gene signature. According to the results obtained by RT-PCR, Kaplan-Meier estimates for RFS and OS in the good and poor profiles patients’ groups were as follows:

- RFS for Good vs. Poor prognosis profile 70 months after surgery: 85 percent vs. 62 percent.
- OS for Good vs. Poor prognosis profile 70 months after surgery: 97 percent vs. 72 percent.

Univariate and multivariate Cox proportional regression analyses were performed to compute a hazard ratio for the risk groups for both endpoints. Only the lymph node status (hazard ratio, 1.2; 95 percent CI, 1.09 to 1.36) and the gene profile (hazard ratio, 6.3; 95 percent CI, 1.28 to 31.07) proved to be independent prognostic variables for OS. Only the number of positive lymph nodes (≤ 3 versus >3) (hazard ratio, 1.13; 95 percent CI, 1.05 to 1.25) and again the gene profile (hazard ratio, 2.74, 95 percent CI, 1.13 to 6.61) were independent prognostic variables for RFS.

In subgroup analyses, the signature did not predict significantly in lymph node negative patients (many of whom received adjuvant chemotherapy), or in women >52 years of age.

The profile predicted both local and distant relapses in the general population of women with breast cancer. In the poor-prognosis group, most patients survived less than 2 years after relapse, regardless of the site of first relapse. In contrast, patients in the good prognosis group usually had low-risk relapses and survived longer than 2 years after relapse.

This study cannot be considered an independent validation of the MammaPrint assay, since only 60 out of 70 genes were considered, the genes were assessed by a different technology (RT-PCR rather than microarray), and the population was far more heavily treated with adjuvant chemotherapy than previously-tested populations. It therefore did not test a population in whom these results would have a clear implication for therapeutic decisions.

Studies Excluded Upon Complete Review

Eden et al., 2004.⁸¹ This paper was excluded because it did not provide new information on the assays investigated. The gene expression markers identified by van't Veer and colleagues²¹ were compared to both conventional markers and newly constructed indices to predict distant metastases. However, analysis was conducted in the same van't Veer cohort patients, and therefore was not a new validation of the 70-gene signature.

Weigelt et al., 2005.⁸² This paper was excluded because it does not include prognostic information for the investigated assays, although it does provide some useful biologic insights. These authors showed that distant metastases display both the same molecular breast cancer subtype and 70-gene prognosis signature as their primary tumors. These results suggest that the capacity to metastasize is an inherent feature of most breast cancers, implying that poor-prognosis breast carcinomas, as classified by the intrinsic gene set or the 70-gene profile, represent distinct disease entities. These findings support the hypothesis that molecular subtypes might originate from different cell types within the breast, therefore reflecting different biological entities and maintained throughout the multistep metastatic process. Indeed the metastatic nature of poor-prognosis breast carcinomas, which are depicted by the 70-gene profile or the luminal B, HER-2 positive, or basal-like molecular subtype, is an inherent feature of breast cancers that remain stable with time and across distinct tumor outgrowth locations within the same individual.

Nuyten et al., 2006.⁸³ This paper was excluded because the authors used a subset of the van de Vijver²⁵ data set and looked at local recurrence.

This group searched for gene expression signatures that predict the risk of local recurrence after breast-conserving therapy (BCT) in a series of 161 early-stage breast cancer patients who were a subset of the original van de Vijver²⁵ cohort. The 70-gene signature, originally designed to predict metastasis, failed to predict local recurrence after BCT.

In this paper other gene signatures were evaluated. The supervised wound-response

signature^{22,84} is the only gene expression profile that could predict a local recurrence after BCT, while both the 70-gene and the primary hypoxia signatures⁸⁵ failed to predict metastases.

Naderi et al., 2007.⁸⁶ This study was excluded because it was not related to the assays investigated for this review. The authors developed a Cox-ranked 70-gene signature, which is a 'new' signature, and it is not related to the MammaPrint test.

Sun et al., 2007.⁸⁷ This paper was also excluded because it is not related to the assays investigated for this review. The author developed a new predictor (with only 3 genes from the 70-gene profile) for recurrence based on the van't Veer data set and used the 70-gene signature for comparison: the new signature performed better than 70-gene signature.

Table 1. Description of the three gene expression profile assays

| Assay General Information | Measurements | Assay procedures |
|--|--|--|
| <p>Assay: Oncotype DX™, Genomic Health</p> <p>Analytic studies: Cronin 2004,⁴⁴ Cronin 2007⁴⁵</p> <p>Clinical validity and utility studies: Chang 2007,⁵⁵ Cobleigh 2005,⁴⁷ Esteva 2005,⁴⁸ Gianni 2005,⁴⁹ Habel 2006,⁵⁰ Mina 2006,⁵¹ Oratz in press,⁵⁶ Paik 2004²⁸, Paik 2006⁵³</p> <p>Economics studies: Lyman 2007,⁷⁵ Hornberger 2005⁶⁷</p> <p>What is measured: 16 cancer genes, 5 normalizing genes by real time RT-PCR</p> <p>To whom it is offered:</p> <ol style="list-style-type: none"> 1) Stage I, ER positive, who will be treated with tamoxifen; 2) Stage II, ER positive, LN negative, who will be treated with tamoxifen; <p>Web site: http://www.genomichealth.com/oncotype/default.aspx</p> | <p>Genes for normalization: ACTB, GAPDH, RPLPO, GUS, TFRC:</p> <p>Cancer related genes (the following functional groups are used to assess patients' Recurrence Score):</p> <p>Proliferation: Ki67, STK15, Survivin, CCNB1, MYBL2</p> <p>HER2: GRB7, HER2</p> <p>Estrogen: ER, PGR, BCL2, SCUBE2</p> <p>Invasion: MMP11, CTSL2</p> <p>Single genes: GSTM1, CD68, BAG1</p> <p>Algorithm: The recurrence score (RS) is obtained in four steps as follows:</p> <ol style="list-style-type: none"> 1. The expression for each gene is normalized relative to the expression of the 5 reference genes. Reference-normalized measurements range from 0 to 15, with a 1-unit increase reflecting approximately a doubling of RNA; 2. Scores for the groups of genes are calculated from individual expression measurements, as follows: $HER2 \text{ group} = 0.9 * GRB7 + 0.1 * HER2$, (set to 8, if less); $ER \text{ group} = (0.8 * ER + 1.2 * PGR + BCL2 + SCUBE2) \div 4$; $Proliferation \text{ group} = (Survivin + Ki67 + MYBL2 + CCNB1 + STK15) \div 5$ (set to 6.5, if less); $Invasion \text{ group} = (CTSL2 + MMP11) \div 2$ 3. The unscaled recurrence score (uRS) is calculated, using predefined coefficients defined in the three training sets: $uRS = +0.47 * HER2 \text{ group} - 0.34 * ER \text{ group} + 1.04 * proliferation \text{ group} + 0.10 * invasion \text{ group} + 0.05 * CD68 - 0.08 * GSTM1 - 0.07 * BAG1$ 4. The RS is rescaled from the uRS, as follows: $RS = 0$ if $uRS < 0$ $RS = 20 * (uRS - 6.7)$ if $0 \leq uRS \leq 100$ $RS = 100$ if $uRS > 100$ 5. Risk groups: Low risk: $RS \leq 17$ Intermediate risk: $18 \leq RS \leq 30$ High risk: $RS \geq 31$ | <ol style="list-style-type: none"> 1. FFPE specimen shipment. 2. Central pathological review for tumor content. 3. RNA preparation and quality evaluation. 4. Triplicate real time RT-PCR. 5. Gene expression measures normalization. 6. RS score computation and re-scaling; Risk group assignment. <p>Results: Report with RS and Risk Group.</p> |

Table 1. Description of the three gene expression profile assays (continued)

| Assay General Information | Measurements | Assay procedures |
|--|--|--|
| <p>Assay: MammaPrint®, Agendia</p> <p>Analytic studies: Ach 2007,⁵⁷ Glas 2006,⁵⁸</p> <p>Clinical validity and utility studies: Buyse 2006,⁵⁹ van de Vijver 2002,²⁵, van't Veer 2002,²¹</p> <p>Other studies: Fan 2006,⁷⁹ and Espinosa 2005⁸⁰</p> <p>What is measured: gene expression of 1900 genes, including the 70 genes in triplicate, by two-color microarray (Agilent Technologies);</p> <p>To whom it is offered:</p> <ol style="list-style-type: none"> 1) Stage I, ER positive or negative, < 61 years; 2) Stage II, ER positive or negative, LN negative, < 61 years; <p>Web site: http://www.agendia.com/en/Professional/About-MammaPrint/About-MammaPrint</p> | <p>Genes for normalization: ~1800;</p> <p>70-gene signatures (the following functional groups are NOT used to assess patients' risk):</p> <p>Cell signaling, growth factors, transcription: MS4A7, GPR180, RTN4RL1, ZNF533, GPR126, ECT2, ESM1, FGF18, FLT1, GNAZ, STK32B, IGFBP5, IGFBP5, MELK, EBF4, NMU, CDC42BPA, TGFB3, WISP1, <u>SCUBE2</u>;</p> <p>Cell cycle, chromatin, nuclear proteins: TSPYL5, CCNE2, CENPA, CDCA7, LGP2, EXT1, NDC80, MTDH, DTL, NUSAP1, MCM6, ORC6L, PRC1, RFC4;</p> <p>Cell adhesion, cell motility, cytoskeleton organization: AYTL2, DIAPH3, COL4A2, DIAPH3, DIAPH3, MMP9;</p> <p>Metabolism, intracellular transport, Golgi: ALDH4A1, AP2B1, QSOX2, GMPS, GSTM3, PITRM1, OXCT1, Peci, Peci, RAB6A, SLC2A3, EGLN1;</p> <p>Ubiquitination: FBXO31, UCHL5;</p> <p>Apoptosis: BBC3;</p> <p>Drug resistance: DCK;</p> <p>Unknown function: LOC286052, PALM2-AKAP2, AA834945, LOC643008, AA404325, AI283268, AI224578, RUNDC1, C9orf30, AW014921, C16orf61, C20orf46, HRASLS, SERF1A</p> <p>Algorithm: classification of patients into risk groups is obtained as follows:</p> <ol style="list-style-type: none"> 1. Scanned image analysis, background correction for non-specific hybridization, and gene expression measurements normalization. Expression values are expressed as log10 ratios. Probes are excluded from further calculations if their background corrected intensities are below zero and/or if spots are flagged as outliers, as determined by the image analysis software. The overall expression value for each signature gene is computed using an error weighted mean over the triplicate probes. 2. The gene signature risk classification is given as a dichotomized value only: high or low risk. A tumor is defined as having a low-risk gene signature if the cosine correlation coefficient for the expression of the 70-gene profile in that tumor with the previously established classifier is above 0.4, the cut point used in the original study by van't Veer, 2002.²¹ | <ol style="list-style-type: none"> 1. Fresh specimen shipment; 2. Central pathological review for tumor content; 3. RNA preparation and quality evaluation; 4. Dye swap microarray hybridization on MammaPrint®; 5. Image analysis; 6. Gene expression measures normalization; 7. 70-gene signature computation and correlation to 70-gene profile; Risk group assignment; <p>Results: Report with Risk Group;</p> |

Table 1. Description of the three gene expression profile assays (continued)

| Assay General Information | Measurements | Assay procedures |
|--|---|--|
| <p>Assay: Breast cancer Profiling (BCP), also known as H/I assay, AviarADX/ Quest Diagnostics</p> <p>Clinical validity and utility studies: Goetz 2006,⁶² Jansen 2007,⁷² Jerevall 2007,⁶³ Ma 2004,⁶⁴ Ma 2006,⁶¹ Reid 2005,⁶⁹ and Fan 2006⁷⁹</p> <p>What is measured: 6 genes by real time RT-PCR;</p> <p>To whom it is offered: 1) Treatment-naïve individuals with ER positive, LN negative breast cancer;</p> <p>Web site: http://www.aviaradx.com/index.html http://www.questdiagnostics.com/index.html</p> | <p>Genes for normalization: ACTB, HMBS, SDHA, and UBC;</p> <p>Two gene ratio: HOXB17, IL17BR;</p> <p>Algorithm: The H/I index is obtained in four steps as follows:</p> <ol style="list-style-type: none"> 1. For each sample, cycle threshold for the 4 reference genes are averaged to obtain the reference CT (refCT); 2. Relative expression levels for <i>HOXB13</i> and <i>IL17BR</i> are computed as individual CT differences from the refCT; 3. Normalized CT differences are z-transformed and the H/I index is obtained by subtracting these two values; 4. An optimized cut-off point is used to dichotomize patients into two risk groups, low and high | <ol style="list-style-type: none"> 1. FFPE Specimen shipment; 2. Central pathological review for tumor content; 3. Specimen macro-dissection, if needed, LCM; 4. RNA preparation and quality evaluation; 5. Duplicate real time RT-PCR; 6. Gene expression measures normalization and Z-transformation; 7. Two-gene index computation; Risk group assignment; <p>Results: Report with risk group</p> |

FFPE = formalin-fixed paraffin-embedded; RT-PCR = reverse transcriptase polymerase chain reaction; ER = estrogen receptor; LN = lymph node; HER = human epidermal growth factor receptor; RS = recurrence score; CT = cycle threshold; LCM = laser-capture micro dissected

Table 2. Successful assays, Oncotype DX™

| Study, year | Protocol | Measure | Success | Tumor < 5% | Poor RNA | Pathological review | RT-PCR | Low reference genes | Clinically ineligible |
|-------------------------------|----------|---------|-----------------|---------------------|--------------------|---------------------|---------------|---------------------|-----------------------|
| Chang, 2007 ⁵⁵ | Standard | FFPE | 80/97 (82.4%) | 16 /97 (16.5%) | 1/80 (1.2%) | | | | |
| Cobleigh, 2005 ⁴⁷ | Standard | FFPE | 78/85 (91.7%) | 7 (8.2%) | | | | | 1* |
| Esteva, 2005 ⁴⁸ | Standard | FFPE | 149/220 (67.7%) | 42/220 (19.0%)* | 4/220 (1.8%) | | 3 /220 (1.3%) | 22/220 (10%) | |
| Gianni, 2005 ^{49†} | Standard | FFPE | 89/95 (93.7%) | 2 patients (2.1%) | 4 patients (4.2%); | | | | |
| Habel, 2006 ⁵⁰ | Standard | FFPE | 865/790 | 19 (7.9%)‡ | | 59§ | 1%‡ | | |
| Mina, 2006 ^{51†} | Standard | FFPE | 45/57 (78.9%) | 3/57 (5.3%) (< 20%) | 9/57 (15.8%) | | | | |
| Oratz, in press ⁵⁶ | Standard | FFPE | 72 /74 (97.3%) | 2 /74 (2.7%) | | | | | |
| Paik, 2004 ²⁸ | Standard | FFPE | 668/675 (98.9%) | 675/754 (89.5%)¶ | | | | | |
| Paik, 2006 ⁵³ | Standard | FFPE | 651/670 (97.2%) | | | | | | |

* Clinically Ineligible (Stage IV)

† Studies performed according to Oncotype DX protocols, but not performed at Genomic Health

‡ It is not clear what the total number was, since it was not reported

§ 59 tumors underwent macro-dissection; it is not clear whether they were used in the subsequent analyses

|| Out of the 70 eligible patients 57 were analyzed (No consent: 3 patients; No specimen available: 10 patients)

¶ The total of 754 (79+675) patients was computed and not clearly reported in the study

FFPE = formalin fixed paraffin embedded; RNA= ribonucleic acid; RT-PCR= reverse transcriptase polymerase chain reaction.

Table 3. Variability and reproducibility, Oncotype DX™

| Study, year | Aims and Methods: | Results | Conclusions |
|----------------------------|--|---|---|
| Cronin, 2007 ⁴⁵ | <p>To assess individual gene and RS reproducibility</p> <p>Repeated measurements of 2 aliquots of a single RNA across multiple days, operators, RT-PCR plates, 7900HT instruments, and liquid-handling robots</p> <p>Mixed-effect ANOVA was used to estimate components of variance;</p> | <p>Reproducibility, CT measurements SD for the individual genes: Total SD range: 0.06 to 0.15 CT units Between day SD range: 0 to 0.055 CT units Between plated SD range: 0 to 0.090 CT units Within plated SD range: 0.057 to 0.147 CT units</p> <p>At a CT of 30 a maximum SD of 0.15 translates into a CV of 0.5%</p> <p>The largest differences between operators, liquid handling robots, and 7900HT instruments < 0.5 CT</p> <p>Reproducibility, CT measurements SD for the RS: Total SD was 0.792 RS units Between day SD was 0 RS units Between plated was 0 RS units Within plated was 0.792 RS units</p> | <p>Authors reported that the following procedures were performed to assure the reproducibility of the assay: A standard RNA control sample is assayed at least once per batch of patient (46 samples) PCR controls are run in every assay plate RT-PCR failures are excluded from analysis Expression values are assigned when at least 2 of 3 assay wells provide acceptable RT-PCR results All 21 genes must have an expression value assigned for an RS to be calculated and reported</p> |
| Habel, 2006 ⁵⁰ | <p>To assess RS reproducibility</p> <p>Pearson's correlation and ANOVA to assess within-patients correlation and variability: 60 blocks that did not undergo macro-dissection from 20 patients (2 to 5 blocks per patient); 49 core biopsies or tumor resection blocks</p> | <p>RS (as a continuous value) SD and Pearson's correlation observed in two unpublished studies: Overall between blocks SD was 3.0 RS units For 16 of the 20 patients, the between blocks SD was < 2.5 RS units Pearson's correlation = 0.86</p> <p>Similar results from the second study</p> | |
| Paik, 2004 ²⁸ | <p>To assess individual genes and RS reproducibility</p> <p>Reproducibility within and between blocks was assessed by performing the assay in five serial sections from six blocks in two patients</p> | <p>Reproducibility evaluation: 16 Cancer genes SD ranged from 0.07 to 0.21 CT units; Within-block RS SD = 0.72 RS unit (95% CI = 0.55 to 1.04); Total within-patient SD (including between and within-block SD) = 2.2 RS units;</p> <p>Similar variability in the RS was observed in reanalysis of clinical trial samples on separate days with different reagent lots (data not shown).</p> | |

RS = recurrence score; RNA=ribonucleic acid; RT-PCR = reverse transcriptase polymerase chain reaction; ANOVA=analysis of variance; CT = cycle threshold; SD = standard deviation

Table 4. Analytic validity, OncoType DX™

| Study, year | Context and Methods | Results | Comments |
|----------------------------|---|---|---|
| Cronin, 2004 ⁴⁴ | <p>RT-PCR protocol optimization from FFPE specimens</p> <p>Pearson's correlation to compare FFPE and frozen specimen</p> <p>Comparison between FFPE and frozen specimens, 48 genes</p> | Correlation = 0.91, p value < 0.0001 | Gene expression profiling analysis is possible using FFPE blocks and comparable to results obtained from frozen specimens. |
| Cronin, 2007 ⁴⁵ | <p>Analytic components addressed: LOD LOQ Amplification efficiency Dynamic range linearity Accuracy and precision Reproducibility</p> <p>A pooled reference RNA was used to perform repeated measurement and a serial dilutions experiment to assess LOD, LOQ, amplification efficiency, dynamic range linearity, and accuracy and precision</p> <p>In all the experiments template input was equivalent to 2 ng RNA; In the linearity study input varied between 2x10⁻¹⁰ and 2000 ng per reaction</p> <p>Repeated measurements of 2 aliquots of a single RNA sample across multiple days, operators, RT-PCR plates, RT-PCR instruments, and liquid-handling robots were performed to assess individual gene and RS reproducibility by ANOVA</p> | <p>The LOD and the LOQ for all the genes proved to be within the pre-specified limits of CT units (< 40 cycles)</p> <p>Amplification efficiencies for all the genes ranged from 75.3% (GAPDH) to 112.1% (BAG1), with values exceeding 100% due to the cumulative nature of the analysis along the sample-dilution series</p> <p>Assay linearity and dynamic range: 6 genes were linear over the entire range; 4 genes were linear over a range of 2e-8 to 2000 ng; The estimated maximal deviation from linearity was below 1 CT over a linear range > 2000-fold, as specified by CLSI</p> <p>Assay quantitative bias and precision at the 2-ng/well for the 16 cancer-related genes: range = -10% (BAG1) to 6% (CTSL2) estimated mean deviation = 0.3% CV of 5.7%</p> <p>Assay quantitative bias and precision at the 2-ng/well for the 5 reference genes: range = -1.5% (GUSB) to 3.3% (ACTB) estimated mean deviation = 0.7% CV of 3.2%</p> | <p>Authors reported that the following procedures were performed to assure the reproducibility of the assay: A standard RNA control sample is assayed at least once per batch of patient (46 samples) PCR controls are run in every assay plate RT-PCR failures are excluded from analysis Expression values are assigned when at least 2 of 3 assay wells provide acceptable RT-PCR results</p> <p>All 21 genes must have an expression value assigned for an RS to be calculated and reported</p> |

RT-PCR = reverse transcriptase polymerase chain reaction; FFPE = formalin-fixed paraffin-embedded; LOD = limits of detection; LOQ = limits of quantification; RNA=ribonucleic acid; RS = recurrence score; ANOVA = analysis of variance; CT=cycle threshold; CLSI = Clinical and Laboratory Standards Institute; CV=coefficients of variation

Table 5. RT-PCR vs IHC comparison assays, Oncotype DX™

| Study, year | Comparison | Method | Details | Results | Comments |
|------------------------------|--|---|--|--|---|
| Chang, 2007 ⁵⁵ | IHC and RT-PCR Concordance Methods for IHC not described | Cohen's κ statistics | Positivity from RT-PCR: ER, >6.5.0 CT PR, > 5.5 CT HER2, > 11.5 CT | ER, $\kappa = 1.00$, 95% CI = 1.00 to 1.00; PR, $\kappa = 0.57$, 95% CI = 0.37 to 0.77; HER2, $\kappa = 0.74$; 95% CI = 0.45 to 1.00 | The RT-PCR technology provides a potential platform for a predictive test using small amounts of routinely processed specimens (core biopsies) |
| Cobleigh, 2005 ⁴⁷ | IHC and RT-PCR Concordance IHC by standard biotin-streptavidin method and appropriate antibody (DAKO, CA, USA) | Cohen's κ statistics | Positivity from RT-PCR: ER, >6.5.0 CT ER+ if IHC staining was present in more than 10% of cells | ER, $\kappa = 0.83$ PR, $\kappa = 0.40$ HER2, = 0.67 Ki-67/MIB1, = 0.22 | The accuracy and specificity of this RT-PCR assay of formalin-fixed, paraffin embedded tumor tissue was supported by comparison of the results of RT-PCR assay of RNA and IHC assay of protein for ER, PR, and HER2 |
| Cronin, 2004 ⁴⁴ | Methods for IHC and FISH not described | Percentage of agreement | ER positivity from RT-PCR: ER, > 8 CT | ER, 93.5% PR, 84% HER2, 100% | |
| Esteva, 2005 ⁴⁸ | IHC and RT-PCR Concordance IHC methods in Esteva et al., 2003 ⁸⁸ and Wang et al., 2002 ⁸⁹ | Cohen's κ statistics Logistic model (IHC as a quantal response) | Positivity from RT-PCR: ER, >7.0 CT PR, > 6.0 CT HER2, > 11.5 CT RT-PCR specificity and sensitivity, in comparison to IHC for HER2 were obtained at the different thresholds | ER, $\kappa = 0.80$, (=+ 0.05) PR, $\kappa = 0.48$ HER2, $\kappa = 0.6$, (=+ 0.08) Logistic model p value: < 0.001 HER2, RT-PCR > 11.50CT: Specificity: 77% Sensitivity: 84% HER2, RT-PCR > 11.5 CT: Specificity: 89% Sensitivity: 84% HER2, RT-PCR > 12.0 CT: Specificity: 95% Sensitivity: 68% | |

Table 5. RT-PCR vs IHC comparison assays, Oncotype DX (continued)

| Study, year | Comparison | Method | Details | Results | Comments |
|----------------------------|---|---|---|---|--|
| Gianni, 2005 ⁴⁹ | IHC and RT-PCR Concordance IHC by reagents from Lab Vision-Neomarkers (Fremont, CA) | Cohen's κ statistics | Positivity from RT-PCR: ER, >6.5.0 CT ER+ if IHC staining was present in more than 10% of cells | ER, κ = 0.83 PR, κ = 0.40 HER2, = 0.67 Ki-67/MIB1, = 0.22 | |
| Habel, 2006 ⁵⁰ | ER status in medical and RT-PCR Concordance ER status methods not defined | Cohen's κ statistics | Positivity from RT-PCR: ER, >6.5.0 CT | ER = 0.49, 95% CI 0.41–0.56 115/122 discordances ER+ by RT-PCR | |
| Mina, 2006 ⁵¹ | IHC and RT-PCR Concordance IHC by standard biotin-streptavidin method and appropriate antibody (DAKO, CA, USA) | Percentage of agreement | Positivity from RT-PCR: ER, >6.5.0 CT ER+ if IHC staining was present in more than 10% of cells; | 41/45 concordant samples: 2 ER+ by IHC were ER- by RT-PCR 2 ER+ by RT-PCR were ER-by IHC Agreement data for PR, KI-67 and HER2/neu were not reported | Gene expression analysis on core biopsy samples is feasible; |
| Paik, 2004 ²⁸ | ER and PR receptors proteins were measured by ligand-binding assays | HER2 DNA was measured by a fluorescence in situ hybridization assay (PathVysion, Vysis) | | | |

IHC = immunohistochemistry; RT-PCR = real time polymerase chain reaction; ER = estrogen receptor; PR = progesterone receptor; HER-2 = human epidermal growth factor receptor 2; FISH = Fluorescence *in situ* Hybridization; CT = cycle threshold; CI= confidence interval.

Table 6: Successful assays, MammaPrint®

| Study, year | Protocol | Measure | Success | Tumor < 5% | Poor RNA | Pathological review | RT-PCR | Low reference genes | Clinically ineligible |
|---------------------------|-----------------|----------------|-----------------|----------------------|-----------------|----------------------------|---------------|----------------------------|------------------------------|
| Buyse, 2006 ⁵⁹ | MammaPrint | Fresh tumors | 326/403 (80.9%) | | 77/403 (19.1%) | | | | |

RT-PCR=reverse transcriptase polymerase chain reaction.

Table 7. Reproducibility, MammaPrint®

| Study, year | Context and Methods | Results | Conclusions |
|--------------------------|--|--|--|
| Ach, 2007 ⁵⁷ | <p>MammaPrint assay intra- and inter-laboratory reproducibility</p> <p>Variation in RNA amplification and labeling, hybridization and wash, and slide scanning was measured on 4 tumors, dye-swap design, 24 slides (8 per site)</p> <p>Methods: To assess reproducibility in this study, ANOVA P values and Pearson's correlation were used</p> | <p>Replicate hybridizations Pearson's correlation at the same site:</p> <p>For 1 tumor in 1 sub-array = 0.983 For 2 tumor in 2 sub-arrays = 0.988 For all the other technical replicates > 0.993</p> <p>Scanning reproducibility across sites: Cy3: Pearson correlation >0.995, slope = 0.97 Cy5: less reproducible (data not shown)</p> <p>70-gene signature reproducibility: No differences between hybridization sites No differences between hybridization days (regardless of site)</p> <p>Statistically significant difference (P value <0.05) between labeling sites for two tumors</p> | <p>The authors showed very low influence on sample-to-reference ratios based on averaged triplicate measurements in the two-color experiments</p> <p>RNA labeling was the largest contributor to inter-laboratory variation</p> <p>Overall, despite this variation, measurement of 70-gene signature in three different laboratories was found to be highly robust</p> |
| Glas, 2006 ⁵⁸ | <p>Pearson's correlation to assess correlation with original data and reproducibility</p> <p>ANOVA analysis to model variability in repeated experiments using the 70 genes of the signature</p> <p>Reproducibility in time was assessed by repeated measurements of RNA aliquots:</p> <ul style="list-style-type: none"> 1 patient with cosine correlation to Good profile = 0.61, for 12 months 1 patient with cosine correlation to Good profile = -0.44, for 12 months 1 border-line with cosine correlation to Good profile = 0.43, for 4 months | <p>Comparison to original 70-gene signature data, Pearson's correlations and in repeated measurements:</p> <p>78 van't Veer²¹ patients, $r = 0.92$, p value < 0.0001 145 (84+61) van de Vijver²⁵ LN-negative patients: $r = 0.88$ p value < .0001 49 patients analyzed twice, $r = 0.995$</p> <p>Reproducibility results from ANOVA analysis: No variation within individuals (p value = 0.96) Significant variation between individuals and genes</p> <p>Reproducibility in time analysis results: For both patients assessed over a period of 12 months measurements SD was 0.028 of the cosine correlation For the 1 border-line sample assessed over a period of 4 months measurements SD was 0.027 of the cosine correlation This latter sample was misclassified 6 times (15%)</p> | <p>Microarray technology can be used as a reliable diagnostic tool</p> <p>The MammaPrint assay performed similarly to the original 70-genes signature</p> |

ANOVA = analysis of variance; Cy3: the green fluorescent dyes commonly used in two colors design microarray hybridization; Cy5 the red fluorescent dyes commonly used in two colors design microarray hybridization; LN = lymph node; SD = standard deviation

Table 8: Analytic validity, MammaPrint®

| Study, year | Context and Methods: | Results | Comments |
|--------------------------|---|--|---|
| Ach, 2007 ⁵⁷ | <p>Context: MammaPrint assay was used to evaluate the intra- and inter-laboratory reproducibility of the assay involving three laboratories. Variation in RNA amplification and labeling, hybridization and wash, and slide scanning was measured on 4 tumors, dye-flip design, 24 slides (8 per site).</p> <p>Methods: To assess reproducibility in this study, ANOVA P values and Pearson's correlation were used.</p> | <p>Replicate hybridizations Pearson's correlation at the same site: For 1 tumor in 1 sub-array = 0.983 For 2 tumor in 2 sub-arrays = 0.988 For all the other technical replicates > 0.993</p> <p>Scanning reproducibility across sites: Cy3: Pearson correlation >0.995, slope = 0.97 Cy5: less reproducible (data not shown)</p> <p>70-gene signature reproducibility: No differences between hybridization sites No differences between hybridization days (regardless of site)</p> <p>Statistically significant difference (P value <0.05) between labeling sites for two tumors</p> | <p>The authors showed very low influence on sample-to-reference ratios based on averaged triplicate measurements in the two-color experiments;</p> <p>RNA labeling was the largest contributor to inter-laboratory variation;</p> <p>Overall, despite this variation, measurement of 70-gene signature in three different laboratories was found to be highly robust;</p> |
| Glas, 2006 ⁵⁸ | <p>Context: MammaPrint assay development through re-analysis of patients from the van't Veer²¹ and van de Vijver²⁵</p> <p>A different reference RNA was used, as well as a different quantification method</p> <p>Methods: 162 total samples from fresh-frozen specimens:</p> <p>84 patients from the van de Vijver²⁵ cohort</p> <p>All 78 patients form the van't Veer cohort²¹</p> <p>A combination of the two population above: 145 (84+61) LN-negative patients</p> <p>49 patients analyzed twice</p> | <p>Comparison to original 70-gene signature data, Pearson's correlations and in repeated measurements: 78 van't Veer²¹ patients, r = 0.92, p value < 0.0001 145 (84+61) van de Vijver²⁵ LN-negative patients: r = 0.88 p value < .0001 49 patients analyzed twice, r = 0.995</p> <p>Reproducibility results from ANOVA analysis: No variation within individuals (p value = 0.96) Significant variation between individuals and genes</p> <p>Reproducibility in time analysis results: For both patients assessed over a period of 12 months measurements SD was 0.028 of the cosine correlation For the 1 border-line sample assessed over a period of 4 months measurements SD was 0.027 of the cosine correlation</p> <p>This latter sample was mis-classified 6 times (15%)</p> | <p>The authors demonstrate for the first time that microarray technology can be used as a reliable diagnostic tool;</p> <p>The MammaPrint assay performed similarly to the original 70-genes signature</p> |

ANOVA = analysis of variance; Cy3: the green fluorescent dyes commonly used in two colors design microarray hybridization; Cy5 the red fluorescent dyes commonly used in two colors design microarray hybridization; LN = lymph node; SD = standard deviation

Table 9: Successful assays, two-gene signature and H/I ratio assays

| Study, year | Protocol | Measure | Success | Insufficient Tumor | Poor RNA | Pathological review | RT-PCR | Low reference genes | Clinically ineligible |
|------------------------------|-----------------|----------------|----------------------------|-------------------------------|-----------------|----------------------------|---------------|----------------------------|------------------------------|
| Goetz, 2006 ⁶² | Study Specific | FFPE | 206/211 (97.6%)* | 211/227 (93%) | | | 5/211 (2.4%) | | |
| Jerevall, 2007 ⁶³ | Study Specific | FFPE | 357/373 (95.7%%) | | 16/373 (4.3%) | | | | |
| Ma, 2006 ⁶¹ | H/I ratio assay | | 852/870 (98%) [†] | 132/1002 (13.2%) [‡] | 18/870 (2%) | | | | |

* Out of the 256 eligible patients 227 were analyzed (No specimen available: 29 patients)

[†] Out of the 1002 eligible

[‡] Tumor content < 10%

RT-PCR = reverse-transcriptase polymerase chain reaction; FFPE = formalin fixed paraffin embedded

Table 10: Reproducibility, two-gene signature and H/I ratio assay

| Study, year | Methods: | Results |
|------------------------------|---|---|
| Jerevall, 2007 ⁶³ | Reproducibility between two institutions, Pearson's correlations, 10 patients | HOXB13:b-actin, $r = 0.96$, $P < 0.001$ IL17BR:b-actin, $r = 0.87$, $P = 0.002$ HOXB13:IL17BR, $r = 0.99$, $P < 0.001$ |
| Ma, 2004 ⁶⁴ | Correlations between microarray and RT-PCR: 59 patients | HOXB13, $r = 0.83$ IL17BR, $r = 0.93$ HOXB13/IL17BR, $r = 0.83$ |

RT-PCR = real time polymerase chain reaction

Table 11: RT-PCR vs IHC comparison assays, two-gene signature and H/I ratio assays

| Study, year | Comparison | Method | Details | Results | Comments |
|------------------------|---|-----------------------------|---|--|---|
| Ma, 2006 ⁶¹ | IHC and RT-PCR Concordance Methods for IHC in ^{90 91} | Cohen's κ statistics | IHC Allred ⁹² scores of 3 to 8 were considered positive for ER or PR ⁹³ Both ER and PR mRNA RT-PCR measurements were bimodal; midpoints used as cutoffs: 2.5 CT for ER 5.9 CT for PR | ER, 91% concordance, $\kappa = 0.83$, P value = .0001 ER, 85% concordance, $\kappa = 0.70$, P value = .0001 | According to the authors these results confirmed the significant correlations between mRNA and protein levels for ER and PR and provided validation of our FFPE gene expression assay platform. |

IHC = immunohistochemistry; RT-PCR =real time polymerase chain reaction; FFPE = formalin fixed paraffin embedded; ER = estrogen receptor; PR = progesterone receptor; mRNA=messenger ribonucleic acid; CT =cycle threshold

Table 12. Clinical validity, Oncotype DX™

| Study, Population characteristics | End points and Exclusion criteria | Clinical validity and utility results | Conclusions/Comments |
|---|--|---|--|
| <p>Cobleigh, 2005⁴⁷</p> <p>78/86 analyzed patients: mean age 57 y, all LN+; TS 0-2cm: 33%; TS >5cm: 31%; tamoxifen 54%; TG-1: 29%; TG-3: 36%; adjuvant CMF 80%</p> | <p>Metastases; <10 LN+; <5% cancer cells; non invasive breast cancer;</p> <p>10 y DRFS in LN+, ER+ and ER- patients</p> | <p>RS < 18:14% of patients: Recurrence: 29% (95%CI: 0-53%) RS >18 and RS < 31: 24% of patients: 10 years recurrence: 72% (95%CI: 38-88%) RS > 31: 62% of patients 10 years recurrence: 80% (95% CI: 63-89%)</p> | <p>Development of RS</p> <p>GEP was correlated with the likelihood of DRFS</p> |
| <p>Esteva, 2005⁴⁸</p> <p>149/220 analyzed patients: mean age 58 y; ER+: 69.1%, PR+: 66.4%; HER-2+: 16.8%; TG-1: 12.1%, TG-3: 30.2%; median TS 2.3 cm</p> | <p>follow-up < 5y; adjuvant therapy; LN+; <5% of cancer cells;</p> <p>DFRS in LN-, untreated patients;</p> | <p>No significant correlation between age, tumor size, or RS and DFRS</p> <p>No significant difference between RS risk groups with respect to distant recurrence-free survival;</p> | <p>Differences between the NSABP B-14 population used in Paik et al.: ER+ and ER- patients were used; ER+ patients not treated with tamoxifen; Association between high nuclear grade and improved outcome; Patients from a single institution</p> |
| <p>Habel, 2006⁵⁰</p> <p>220/234 analyzed cases: TG-Well: 11%; TG-Poor: 47%;TS < 2: 64%; TS >2cm: 36%;ER+: 76%</p> <p>570/631 analyzed controls: TG-Well: 31%; TG-Poor: 23%; TS < 2: 79%; TS >2cm: 21%;ER+: 90%</p> | <p>LN+, age >75 y; initially treated with chemotherapy; metastases, inflammatory or bilateral breast cancer; unknown tamoxifen; prior cancer;</p> <p>10 y breast cancer-specific mortality in ER+, LN patients;</p> | <p>RS <18 / ER+, tamoxifen: 2.8%, 95%CI: 1.7-3.9 RS <18 / ER+, no tamoxifen: 6.2%, 95%CI: 4.5-7.9 RS 18-30 / ER+, tamoxifen: 10.7%, 95%CI: 6.3-14.9 RS 18-30 / ER+, no tamoxifen: 17.8%, 95%CI: 11.8-23.3 RS ≥31 / ER+, tamoxifen: 15.5%, 95%CI: 7.6-22.8 RS ≥31 / ER+, no tamoxifen: 19.9%, 95%CI: 14.2-25.2</p> | <p>RS associated with risk of breast cancer death in: ER+ patients treated with tamoxifen; ER+ patients not treated with tamoxifen; ER- patients; Such associations remained after accounting for tumor size and grade. Moreover the RS was able to identify a larger subset of patients with low risk of breast cancer death than it was possible with either of these standard prognostic indicators</p> |

Table 12. Clinical validity, Oncotype DX (continued)

| Study, Population characteristics | End points and Exclusion criteria | Clinical validity and utility results | Conclusions/Comments |
|--|---|---|--|
| Paik, 2004 ²⁸ 668/754 analyzed patients (tamoxifen treatment arm of NSABP B-14) | <5% cancer cells; insufficient RNA; weak RT-PCR signal (average cycle threshold for reference genes >35); distant recurrence and the Overall Survival in LN-, ER+ breast cancer; | 51% of the patients RS <18, KM estimates = 6.8, 95%CI = 4.0–9.6 22% of the patients RS > 18 <31, KM estimates = 14.3-95%CI=8.3–20.3 27% of the patients RS > 31, KM estimates =30.5-95%CI=23.6–37.4) PR, ER, HER, age, size, grade, and RS: p-Value=0.001, Hazard Ratio =2.81 (95%CI= 1.70–4.64 for 50 units increase | RS validated in tamoxifen-treated, LN-, ER+ breast cancer patients |
| Paik, 2006 ⁵³ 651/670 analyzed patients: TS < 2: 66%; TG-Well: 13%; TG-Poor: 28%; TS >2cm: 34%; ER+: 100%, LN-: 100% tamoxifen treatment arm of NSABP B-20 | <5% invasive tumor; insufficient RNA; weak RT-PCR signal (average cycle threshold for reference genes >35); distant recurrence in ER+, LN- breast cancer from NSABP B20 | 20.6% of the patients RS<18, tamoxifen, 96.8 93.7% to 99.9% 33.5% of the patients RS<18, chemotherapy, 95.6 92.7% to 98.6% 7% of the patients RS >18 <31, tamoxifen: 90.9 82.5% to 99.4% 13.7% of the patients RS >18 <31, chemotherapy: 89.1 82.4% to 95.9% 7.2% of the patients RS>31, tamoxifen: 60.5 46.2% to 74.8% 18% of the patients RS>31, chemotherapy: 88.1 82.0% to 94.2% | The RS assay predicts the magnitude of chemotherapy benefit in women with node-negative, ER–positive breast cancer If RS risk groups are considered: a minimal benefit from chemotherapy is seen in the low risk group, however with large intervals benefit is not assessable in the Intermediate risk group due to the uncertainty in the estimates a large chemotherapy benefit is seen in the high risk group |

LN = lymph node; TS = tumor size; TG = tumor grade; CMF = cyclophosphamide, methotrexate, and fluorouracil; DRFS= distant recurrence-free survival; ER = estrogen receptor; CI=confidence interval; RS = recurrence score; GEP=gene expression programming; HER = human epidermal growth factor receptor; NR = not reported; pCR=complete pathological response; INT= Italian National Cancer Institute of Milan, Italy; NSABP = The National Surgical Adjuvant Breast and Bowel Project; RT-PCR = reverse transcriptase polymerase chain reaction; PR = progesterone receptor; KM=Kaplan Meier.

Table 13. Risk classification of Oncotype DX™ against the St. Gallen criteria

| St. Gallen risk group | 10 yr. risk of distant relapse by St. Gallen | Oncotype DX Risk group | 10 yr. risk of distant relapse | % of St. Gallen stratum (n) |
|--------------------------------|---|-------------------------------|---------------------------------------|------------------------------------|
| Low (n=53) 8% | 5% | Low | 0% | 72% (38) |
| | | Medium | 18% | 22% (12) |
| | | High | 43% | 6% (3) |
| Intermediate (n=222) 33% | 9% | Low | 5% | 60% (134) |
| | | Medium | 6% | 23% (51) |
| | | High | 21% | 17% (37) |
| High (n=393) 59% | 18% | Low | 8% | 42% (134) |
| | | Medium | 18% | 22% (51) |
| | | High | 33% | 36% (37) |

Table 14. Risk classification of Oncotype DX™ against the 2004 NCCN guidelines

| 2004 NCCN risk group | 10 yr. risk of distant relapse by NCCN | Oncotype DX Risk group | 10 yr. risk of distant relapse | % of NCCN stratum (n) |
|-----------------------------|---|-------------------------------|---------------------------------------|------------------------------|
| Low (n=53) 8% | 5% | Low | 0% | 72% (38) |
| | | Medium | 18% | 22% (12) |
| | | High | 43% | 6% (3) |
| High (n=615) 92% | 15% | Low | 8% | 49% (300) |
| | | Medium | 14% | 22% (137) |
| | | High | 30% | 29% (178) |

NCCN = National Comprehensive Cancer Network

Table 15. Risk classification of Oncotype DX™ against the Adjuvant! guidelines

| Adjuvant! Online risk group | 10 yr. risk of distant relapse by Adjuvant! | Oncotype DX Risk group | 10 yr. risk of distant relapse | % of Adjuvant! Online stratum (n) |
|------------------------------------|--|-------------------------------|---------------------------------------|--|
| Low (n=354) 53% | 8% | Low | 6% | 61% (216) |
| | | Medium- High | 13% | 39% (138) |
| Med-High (n=314) 47% | 22% | Low | 9% | 39% (122) |
| | | Medium -High | 31% | 61% (192) |

Table 16. Clinical Validity, MammaPrint® and 70-gene signature

| Study, Population characteristics | End points and Exclusion criteria | Clinical validity and utility results | Conclusions/Comments |
|---|---|--|---|
| <p>Buyse, 2006⁵⁹</p> <p>307/403 analyzed patients, all age < 60 y, all < 5cm (ER missing: 5 patients)</p> <p>Clinical low risk/gene low risk n=52 (TS < 2cm: 67%, ER+: 100%, TG-Good: 43%, TG-Poor: 0%)</p> <p>Clinical low risk/gene high risk n=28 (TS < 2cm: 59%, ER+: 100%, TG-Good: 43%, TG-Poor: 0%)</p> <p>Clinical high risk/gene low risk n=59 (TS < 2cm: 29%, ER+: 91%, ER-: 9%, TG-Good: 12%, TG-Poor: 18%)</p> <p>Clinical high risk/gene high risk n=163 (TS < 2cm: 25%, ER+: 48%, ER-: 52%, TG-Good: 3%, TG-Poor: 69%)</p> | <p>Exclusion criteria: Age > 61 y, TS >5cm, previous malignancy (except basal cell carcinoma), bilateral synchronous breast</p> <p>End points: OS, RFS, TTM</p> | <p>KM analysis stratified by MammaPrint and Adjuvant (% of patients with distant recurrence): Good(R>0.4), Adjuvant!Low: 52 patients OS(10years): 0.88 (0.74 to 0.95) Good(R>0.4), Adjuvant!High: 59 patients OS(10years): 0.89 (0.77 to 0.95) Poor(R<0.4), Adjuvant!Low: 28 patients OS(10years): 0.69 (0.45 to 0.84) Poor(R<0.4), Adjuvant!High: 163 patients OS(10years): 0.69 (0.61 to 0.76)</p> <p>Hazard Ratios (unadjusted), MammaPrint: TTM=2.32 (95% CI = 1.35-4.00) DFS=1.50 (95% CI = 1.04-2.16) OS=2.79 (95% CI = 1.60-4.87)</p> <p>MammaPrint adjusted by Adjuvant: TTM= 2.13 (95% CI = 1.19 to 3.82) DFS= 1.36 (95% CI = 0.91 to 2.03) OS= 2.63 (95% CI =1.45 to 4.79)</p> <p>Development of metastases within 5 years: Sensitivity for Gene signature 0.90 (0.78 to 0.95) Sensitivity for Adjuvant! 0.87 (0.75 to 0.94) Specificity for Gene signature 0.42 (0.36 to 0.48) Specificity for Adjuvant! 0.29 (0.24 to 0.35)</p> <p>ROC area under the curve: MammaPrint®: TTM: 0.681 MammaPrint®: OS: 0.659 Adjuvant: TTM: 0.648 Adjuvant: OS: 0.576</p> | <p>MammaPrint is a better predictor of TTM than Age, Size, Grade, ER, Adjuvant!, NPI, St Gallen</p> <p>St Gallen is a better predictor of DFS than MammaPrint</p> <p>MammaPrint is a better predictor for OS than Age, Size, Grade, ER, Adjuvant!, NPI, St Gallen</p> <p>The signature remained a statistically significant prognostic factor for TTM and OS even after adjustment for various risk classifications methods based on clinicopathologic factors</p> <p>The lack of statistical significance for DFS was explained by the fact that the signature was originally developed using TTM as the endpoint</p> <p>Overall the 70-gene signature adds independent prognostic information to clinicopathologic risk assessment for node-negative untreated patients with early breast cancer</p> <p>Clinical risk hazard ratios, adjusted for the gene signature were not significant, suggesting that most of their prognostic utility is subsumed by the gene signature</p> |

Table 16. Clinical Validity, MammaPrint® and 70-gene signature (continued)

| Study, Population characteristics | End points and Exclusion criteria | Clinical validity and utility results | Conclusions/Comments |
|---|---|--|--|
| <p>van't Veer, 2002²¹</p> <p>Population: 78 + 19 patients, (mean age 44.9 y, TS < 2cm: 41.2%, ER+: 70.2%. PR+: 57.7%, LN-: 100%, TG-Good: 12%, TG-Poor: 49%)</p> <p>Therapy: Hormonal (3 patients), Chemotherapy (3 patients)</p> | <p>Exclusion criteria: Age >55 y, TS >5cm, metastases, previous malignancy, diagnosed before 1983, or after 1996</p> <p>End point: distant metastases as first relapse event (5 years)</p> <p>Outcome: No metastases within 5yrs: 51; Metastases within 5yrs: 46</p> | <p>65/78 correct predictions: 5 poor in the 70-gene Good group 8 good in the 70-gene Poor group</p> <p>17/19 correct predictions: 1 poor in the 70-gene Good group 1 good in the 70-gene Poor group</p> <p>Univariate OR=15, (95%CI=4-56, P =0.0000041) Multivariate OR = 18, (95%CI=3.3-94, P = 0.00014)</p> | <p>The 70-genes signatures is a better predictor of the risk of distant metastases than standard clinical predictors</p> |
| <p>van de Vijver, 2002²⁵</p> <p>Population: 295 patients, all age < 53 y, all < 5cm, 61 in common with van't Veer 2002:²¹</p> <p>Poor Prognosis n=180, (TS < 2cm: 647%, LN-: 51%, ER+: 63%); Hormonal (13% patients), Chemotherapy (37% patients)</p> <p>Good Prognosis n=115, (TS < 2cm: 62%, LN-: 52%, ER+: 97%); Hormonal (15% patients), Chemotherapy (38% patients)</p> | <p>Exclusion criteria: Age >52 y, TS >5cm, previous malignancy, apical axillary LN involvement</p> <p>End point: distant metastases as first relapse event, OS</p> | <p>The 70-genes association with age, tumor grade, ER (P value<0.001), and tumor size (P =0.012);</p> <p>67 LN- patients (not in van't Veer 2002²¹) OR = 15.3, (95%CI = 1.8-127, P = 0.003)</p> <p>180 LN+ and LN- patients (not in van't Veer 2002²¹): OR = 14.6, (95%CI = 4.3-50, P < 0.0001)</p> <p>All patients, HR = 5.1, (95%CI = 2.9-9.0, P < 0.001) 151 LN-, HR = 5.5, (95%CI = 2.5-12.2, P < 0.001) 144 LN+, HR = 4.5, (95%CI = 2-10.2, P < 0.001)</p> <p>Multivariate HR = 4.6, 95%CI = 2.3-9.2, p value < 0.001</p> | <p>The authors demonstrate for the first time that microarray technology can be used as a reliable diagnostic tool</p> <p>The MammaPrint assay performed similarly to the original 70-genes signature and is, therefore, an excellent tool to predict outcome of disease in breast cancer patients</p> |

Table 16. Clinical Validity, MammaPrint and 70-gene signature (continued)

| Study, Population characteristics | End points and Exclusion criteria | Clinical validity and utility results | Conclusions/Comments |
|--|--|--|--|
| <p>Glas, 2006⁵⁸</p> <p>Population: 162 LN-, untreated patients (<55 years), from the van de Vijver and van't Veer cohorts</p> <p>All 78 patients from the van't Veer were re-analyzed</p> | <p>Exclusion criteria: See van't Veer, 2002²⁵ above</p> <p>End point: distant metastases as first relapse event</p> | <p>78 patients from the van't Veer²¹ series: MammaPrint OR = 13.95 (95%CI = 3.9-44); 70-genes signature OR = 15, 95%CI = 2.1 to 19) 7/78 differently classified by MammaPrint</p> <p>145 LN- patients from the van de Vijver²⁵ series: MammaPrint HR = 5.6 (95%CI = 2.4-7.3, P = 0.0001)</p> <p>Similar results were obtained for OS</p> | <p>The authors demonstrate for the first time that microarray technology can be used as a reliable diagnostic tool</p> <p>The MammaPrint assay performed similarly to the original 70-genes signature and is, therefore, an excellent tool to predict outcome of disease in breast cancer patients</p> |

ER = estrogen receptor; TS = tumor size; TG = tumor grade; OS=overall survival; RFS=relapse free survival; TTM=time to distant metastases; ROC = Receiver operating characteristic; NPI=Nottingham prognostic index; DFS=disease free survival; OR = odds ratio; CI = confidence interval; LN = lymph node; HR=hazard ratio; KM=Kaplan- Meier.

Table 17. MammaPrint® compared with traditional composite risk markers

| Criteria | Number classified as “high risk” | |
|-------------------------|--|---|
| | Sensitivity Patients with metastases n=34 | Specificity Patients without metastases n=44 |
| St. Gallen ⁵ | 33 (97%) | 31 (30%) |
| NIH ^{3,4} | 32 (94%) | 40 (9%) |
| 70-Gene panel | 31 (91%) | 18 (59%) |

Table 18. Clinical Validity, two-gene signature and H/I ratio assays

| Study, Year | Population size, N | End Points and Major Findings | Comments |
|----------------------------|--|---|--|
| Goetz, 2006 ⁶² | <p>Population: 206/256 eligible patients, from the randomized NCCTG 89-30-52 trial on tamoxifen treatment</p> <p>(TS < 3cm: 76%, LN-: 63%, ER+: 100%, HER2, 0; 11%, HER2, 1; 36%, HER2 2; 34%, TG-1: 26%, TG-3: 18%)</p> <p>Exclusion criteria: NR</p> | <p>End points: RFS (any event of recurrence), DFS (recurrence, or death), and OS (death)</p> <p>Clinical validity and utility results: Nodal status, tumor size and Nottingham grade significantly associated with endpoints</p> <p>All patient in the study: RFS, Multivariate F-S 1.45 (95% CI 0.93, 2.27) DFS, Multivariate F-S 1.57 (95% CI 1.04, 2.38) OS, Multivariate F-S 1.29 (95% CI 0.81, 2.08)</p> <p>Node negative patients in the study (n=130): RFS, Multivariate F-S 1.73 (95% CI 0.92, 3.25) DFS, Multivariate F-S 1.77 (95% CI 0.99, 3.16) OS, Multivariate F-S 2.01 (95% CI 1.02, 3.99)</p> | <p>According to the authors a high 2-gene expression ratio is associated with increased relapse and death in patients with node-negative ER positive breast cancer treated with tamoxifen</p> <p>In this study the 2-gene ration was normalized by standard curve, and no reference genes were used; optimized cut-off points were identified and used</p> |
| Jansen, 2007 ⁷² | <p>Population: 1,252/1693 eligible patients, subsets: DFS: ER+, LN-, no Adjuvant (N = 468) PFS: ER+, first-line tamoxifen (N = 193)</p> <p>Exclusion criteria: distant recurrence within the first month of surgery, missing LN, ER, and HOXB13 and IL17B, < 30% tumor cells in specimens, poor RNA quality</p> | <p>End points: disease-free survival (DFS), progression free survival (PFS), post-relapse survival (PRS), and overall survival (OS)</p> <p>Clinical validity and utility results: Multivariate analysis, ER+, LN-, no Adjuvant (N = 468): Dichotomized ratio with 3' I17RB, DFS HR = 1.74; 95% CI = 1.17 to 2.59; <i>P</i> = 0.006</p> <p>Multivariate analysis, relapsing ER+, tamoxifen (N = 193): Optimal dichotomized ratio with 3' I17RB, PFS HR = 2.97; 95% CI = 1.82 to 4.86; <i>P</i> < 0.001</p> <p>Standard* dichotomized ratio with 3' I17RB, PFS HR = 1.95; 95% CI = 1.39 to 2.73; <i>P</i> < 0.001</p> <p>*as in Ma et 2006⁶¹</p> | <p>High HOXB13-to-IL17BR ratio expression levels associate with both tumor aggressiveness and tamoxifen therapy failure</p> <p>The ratio was significantly associated with DFS and PFS in the specific subsets of patients</p> <p>In multivariate analysis, the ratio was associated with a shorter DFS for node-negative patients only</p> <p>Expression levels normalized to a different set of control genes respect to MA et al 2006⁶¹ using fresh frozen samples</p> |

Table 18. Clinical Validity, two-gene signature and H/I ratio assays (continued)

| Study, Year | Population size, N | End Points and Major Findings | Comments |
|------------------------------|--|---|---|
| Jerevall, 2007 ⁶³ | <p>Population: 357 patients analyzed, 264 post-menopausal, and 93 pre-menopausal.</p> <p>Postmenopausal patients: randomized clinical trial, comparing 2 years (163 patients, 62%) and 5 years (101 patients, 38%) of adjuvant tamoxifen treatment.</p> <p>Exclusion criteria: NR</p> | <p>End points: Correlation with clinical prognostic factors.</p> <p>The ratio was significantly associated to: Tumor size, P = 0.003; ER, P < 0.001; PR, P < 0.001; HER2, P = 0.003; NHG, P < 0.001; Ploidy, P < 0.001; S-phase, P = 0.005; ER, HER2, S-phase and NHG correlations are mostly due to IL17BT; PR and ploidy correlation have contribution from both genes</p> | <p>Lower expression of IL17BR, but not HOXB13, was correlated to several factors related to poor prognosis, IL17BR might be an independent prognostic factor in breast cancer</p> |
| Ma, 2004 ⁶⁴ | <p>Population: by RT-PCR, Frozen specimens: 59/60, FFPE: 20/20 eligible patients</p> <p>Frozen, recurrence, n=28, (mean age: 65.1, LN-: 57%, TS > 5cm: 9, ER+: 97%, TG-1: 7%, TG-3: 39%)</p> <p>Frozen, non recurrence, n=32, (mean age: 69.1, LN-: 47%, ER+: 100%, TG-1: 3%, TG-3: 22%)</p> <p>FFPE, recurrence, n=10, (mean age: 65.5, LN-: 80%, ER+: 100%, TG-1: 10%, TG-3: 30%)</p> <p>FFPE, non recurrence, n=10, (mean age: 65.2, LN-: 100%, ER+: 100%, TG-1: 10%, TG-3: 10%)</p> <p>Exclusion criteria: NR</p> | <p>End points: DFS (months) was calculated from the date of diagnosis.</p> <p>Clinical validity and utility results: Frozen, recurrence, DFS: 54.8 (range: 5-137) Frozen, non recurrence, DFS: 115.6 (range: 61-169) FFPE, recurrence, DFS: 51.4 (range: 15-117) FFPE, non recurrence, DFS: 95.8 (range: 25-123)</p> <p>KM analysis, log-rank test, RT-PCR on the training set: P value = 0.0000058</p> <p>KM survival analysis, log-rank test, RT-PCR on the validation set: P value = 0.002</p> <p>Classification results in the validation set (RT-PCR data): 16/20 correctly classified</p> | <p>The authors concluded that HOXB13/IL17BR 2-gene ratio predicts tumor recurrence in the setting of tamoxifen therapy</p> |

Table 18. Clinical Validity, two-gene signature and H/I ratio assays (continued)

| Study, Year | Population size, N | End Points and Major Findings | Comments |
|--------------------------|---|--|--|
| Ma, 2006 ⁶¹ | <p>Population: 852/1002 eligible patients: All samples, n=852, (age > 50 y: 82%, LN-: 72%, ER+: 73%); Tamoxifen treated, 286, (age > 50 y: 91%, LN-: 40%, ER+: 89%); Untreated, 566, (age > 50 y: 77%, LN-: 84%, ER+: 65%);</p> <p>Exclusion criteria: NR</p> | <p>End points: Relapse-free survival (RFS), defined as the time from initial diagnosis to any recurrence. Optimal threshold for dichotomization of the 2-gene ratio was identified and applied in the analysis</p> <p>Clinical validity and utility results: Two-gene index on a continuous scale, 5-year recurrence risk for untreated patients: Index of -2.0 = 15% (95% CI, 9.8% to 20.5%) Index of +2.0 = 36% (95% CI, 26.5% to 45.2%)</p> <p>Multivariate Cox Regression Analysis; ER+ node negative, untreated test set and tamoxifen treated patients (n = 225), dichotomized HOXB13:IL17BR index (high versus low)</p> <p>HR = 3.9 (95% CI = 1.5 to 10.3) p value = 0.007</p> | <p>According to the authors these results confirmed the significant correlations between mRNA and protein levels for ER and PR and provided validation of the FFPE gene expression assay platform</p> <p>The HOXB13:IL17BR index was only significant in node-negative patients</p> <p>Higher HOXB13:IL17BR index was associated with a higher risk of relapse</p> <p>Two-gene index was a significant predictor of clinical outcome in ER+, node-negative, patients irrespective of tamoxifen therapy</p> |
| Reid, 2005 ⁶⁹ | <p>Population: Tamoxifen 20mg/day for 5 years: 58 patients, (Age > 50yrs: 93.1%, TS ≤2cm: 37.9%, LN-: 22.5%, HER2+: 20.7%, PR+: 79.3%, ER+: 100%)</p> <p>Exclusion criteria: NR</p> | <p>End points: Disease Free Survival (DFS)</p> <p>Clinical validity and utility results: Univariate logistic regression: odds ratio: HOXB13 OR = 1.04, 95% CI = 0.92 to 1.16, P = 0.54 IL17BR OR = 0.69, 95% CI = 0.40 to 1.20, P = 0.18 HOXB13/IL17BR OR = 1.30, 95% CI = 0.88 to 1.93, P = 0.18</p> <p>Similar results by the other methods</p> | <p>Although the proposed predictive model is very appealing the use of the two-gene ratio signature in an independent population yielded statistically non significant results</p> <p>The authors failed to confirm the association of the 2-gene ratio with response to tamoxifen on their cohort (which is however different in terms of clinical characteristics from the original Ma 2004 cohort)</p> <p>The authors also failed to classify patients using Discriminant Linear Analysis on two published data sets, including the Ma 2004 original series</p> |

NCCTG= North Central Cancer Treatment Group; TS = tumor size; LN=lymph node; ER= estrogen receptor; HER-2= Human epidermal growth factor receptor 2; TG = tumor grade; NR = not reported; RFS = relapse-free survival; DFS = disease-free survival; OS = overall survival; CI= confidence interval; RNA=ribonucleic acid; PFS = progression-free survival; PRS = post-relapse survival; NHG= Nottingham histologic grade; PR = progesterone receptor; RT-PCR = reverse transcriptase polymerase chain reaction; FFPE = formalin-fixed paraffin-embedded; KM = Kaplan Meier; HR= hazard ratio.

Table 19. Clinical Utility, Oncotype DX™

| Study, Year | Population size, N | End Points and Major Findings | Comments |
|----------------------------|--|--|---|
| Chang, 2007 ⁵⁵ | <p>Population: 72/97 eligible patients, (mean age 48.5 y, ER+: 69.1%;TG-Well: 2.8%; TG-Poor: 56.9%, LN-: 90%, HER-2+: 13.5%, treated with docetaxel)</p> <p>Exclusion criteria: NR</p> | <p>End points: prediction of clinical response (by the RECIST method) to docetaxel treatment in women with breast cancer</p> <p>Clinical validity and utility results: Clinical CR was more likely in the high RS risk group (P = 0.008); A 50 units increase in the RS showed an OR = 5 (95% CI = 1.3-6.0); AUROC = 0.73;</p> | <p>The authors concluded that Oncotype DX can be potentially be used as a predictive test of chemosensitivity using small amounts of routinely processed specimens</p> |
| Gianni, 2005 ⁴⁹ | <p>Population: 89/95 patients (mean age 49.9 y, stage-T4b: 79%, stage-T4d: 18%, stage-T2: 1%, stage-T3: 2%, ER+: 58%, TG-1: 24%, TG-3: 21%, LN-: 16%, adjuvant with doxorubicin/paclitaxel followed by paclitaxel)</p> <p>Exclusion criteria: NR</p> | <p>End points (goal): to examine the correlation between RS and pCR, and to identify additional genes associated with pCR</p> <p>Clinical validity and utility results: The global likelihood ratio test assessing probit regression based models with and without the incorporation of the RS resulted in a P value of 0.005</p> | <p>The authors showed that the RS was strongly correlated with pCR, and identified a set of genes, whose expression correlated with pCR to neoadjuvant doxorubicin and paclitaxel</p> |
| Habel, 2006 ⁵⁰ | <p>Population: 220/234 eligible cases (TG-Well: 11%, TG-Poor: 47%, TS < 2: 64%, TS >2cm: 36%, ER+: 76%, ER-: 24%)</p> <p>570/631 eligible control patients (TG-Well: 31%, TG-Poor: 23%, TS < 2: 79%: TS >2cm: 21%, ER+: 90%)</p> <p>Exclusion criteria: LN+, age >75 y, initially treated with chemotherapy, inflammatory or bilateral cancer, metastases, prior invasive cancer, unknown/unconfirmed tamoxifen</p> | <p>End points: The risk of breast cancer-specific mortality among women with ER+, LN- breast cancer. Patients were matched by age, race, year of diagnosis and tamoxifen treatment.</p> <p>Clinical validity and utility results: 10 y death risk according to RS (with tumor size and grade): RS <18 / ER+, tamoxifen: 2.8%, 95%CI: 1.7-3.9; RS <18 / ER+, no tamoxifen 6.2%, 95%CI: 4.5-7.9; RS 18-30 / ER+, tamoxifen 10.7%, 95%CI: 6.3-14.9; RS 18-30 / ER+, no tamoxifen 17.8%, 95%CI: 11.8-23.3; RS ≥31 / ER+, tamoxifen 15.5%, 95%CI: 7.6-22.8; RS ≥31 / ER+, no tamoxifen 19.9%, 95%CI: 14.2-25.2;</p> | <p>The authors showed that the RS was strongly associated with risk of breast cancer death among :</p> <ul style="list-style-type: none"> ER+ patients treated with tamoxifen ; ER+ patients not treated with tamoxifen ER- patients <p>Such associations remained after accounting for tumor size and grade. Moreover the RS was able to identify a larger subset of patients with low risk of breast cancer death than it was possible with either of these standard prognostic indicators</p> |

Table 19. Clinical Utility, Oncotype DX (continued)

| Study, Year | Population size, N | End Points and Major Findings | Comments |
|--------------------------|--|---|--|
| Mina, 2006 ⁵¹ | <p>Population: 45/70 eligible patients (mean age 49 y, median TS 6.8 cm, TG-Well: 24%, TG-Poor: 49%, ER+: 57%, HER2+: 18%, LN+: 47%, adjuvant with doxorubicin / docetaxel, tamoxifen in ER+);</p> <p>Exclusion criteria: NR</p> | <p>End points: complete pathological response (pCR) to primary chemotherapy with anthracycline- and taxanes;</p> <p>Clinical validity and utility results: No correlation between Oncotype DX RS and pCR;</p> | Though the Oncotype DX RS correlated with pCR in the INT Milan cohort of the Gianni et al study, ⁴⁹ this association was not found in the present study |
| Paik, 2006 ⁵³ | <p>Population: 651/670 eligible patients (TS < 2: 66%, TG-Well: 13%, TG-Poor: 28%, TS >2cm: 34%, ER+: 100%, LN-: 100%, tamoxifen treatment arm of NSABP B-20)</p> <p>Exclusion criteria: specimen shows <5% invasive tumor, insufficient RNA extracted from specimen, weak RT-PCR signal (average cycle threshold for reference genes >35)</p> | <p>End points: freedom from distant recurrence in women with ER-positive, node-negative breast cancer from NSABP B-20.</p> <p>Clinical validity and utility results: 20.6% of the patients RS<18, Tamoxifen, 96.8 93.7% to 99.9%; 33.5% of the patients RS<18, Chemotherapy, 95.6 92.7% to 98.6%; 7% of the patients RS >18 <31, Tamoxifen: 90.9 82.5% to 99.4%; 13.7% of the patients RS >18 <31, Chemotherapy: 89.1 82.4% to 95.9%; 7.2% of the patients RS>31, Tamoxifen: 60.5 46.2% to 74.8%; 18% of the patients RS>31, Chemotherapy: 88.1 82.0% to 94.2%;</p> | <p>The RS assay predicts the magnitude of chemotherapy benefit in women with node-negative, ER-positive breast cancer;</p> <p>If RS risk groups are considered: a minimal benefit from chemotherapy is seen in the low risk group, however with large intervals; benefit is not assessable in the Intermediate risk group due to the uncertainty in the estimates; a large chemotherapy benefit is seen in the high risk group</p> |

ER = estrogen receptor; TG = tumor grade; LN = lymph node; HER = human epidermal growth factor receptor; NR = not reported; RECIST = response evaluation criteria in solid tumors; CR = complete response; RS = recurrence score; OR = odd ratio; CI = confidence interval; AUROC = area under Receiver operator curve; pCR = pathological complete response; TS = tumor size; RT-PCR = reverse transcriptase polymerase chain reaction; INT = Italian National Cancer Institute of Milan, Italy; NSABP = The National Surgical Adjuvant Breast and Bowel Project.

Table 20. Comparison of economic studies

| Study | Test evaluated | Comparison (guidelines) | Economic outcomes evaluated | Estimated cost difference | Estimated difference in mean QALYs | Confidence in the analysis |
|---------------------------|-----------------------|--------------------------------|---|---|--|-----------------------------------|
| Hornberger ⁶⁷ | Oncotype DX | NCCN | Cost, QALY, DRFS, OS | \$2,028 in favor of RS | 0.086 in favor of RS | Moderate |
| Lyman ⁷⁵ | Oncotype DX | NCCN | Cost, LYS, C/E, ΔCost, ΔLYS, ΔC/E LYS, QALY | RS = \$4,272 vs. Tam RS = -\$2,255 vs. ChemoRx+Tam | RS = +0.97 vs. Tam RS = +1.71 vs. ChemoRx+Tam | Weak |
| Oestreicher ⁷⁶ | GEP | NIH | Cost, QALY | \$2,882 in favor of GEP | 0.22 in favor of NIH | Strong |

QALY = Quality Adjusted Life Years; NCCN = National Comprehensive Cancer Network; DRFS = Distant Recurrence-Free Survival; OS = Overall Survival; RS = Recurrence Score; LYS= Life Years Saved; C/E=Cost Effectiveness; Δ= Change in; ChemoRx= Chemotherapy; Tam=Tamoxifen; GEP = Gene Expression Profiling; NIH = National Institutes of Health.

Table 21. Clinical Utility, two-gene signature and H/I ratio

| Study, Year | Population size, N | End Points and Major Findings | Comments |
|------------------------------|---|---|--|
| Jerevall, 2007 ⁶³ | <p>Population: 357 patients analyzed, 264 post-menopausal, and 93 pre-menopausal.</p> <p>Postmenopausal patients: randomized clinical trial, comparing 2 years (163 patients, 62%) and 5 years (101 patients, 38%) of adjuvant tamoxifen treatment.</p> <p>Exclusion criteria: NR</p> | <p>End points: Relapse-free survival (RFS), defined as the time from diagnosis to local, regional, or distant recurrence or death due to breast cancer; OS, defined as the time elapsed from diagnosis to the date of death due to breast cancer</p> <p>Clinical validity and utility results: Post-menopausal ER+ patients, low ratio: benefit from prolonged tamoxifen (P = 0.021; in KM analysis for RFS) due to the low expression of HOXB13 genes (P = 0.010, in KM analysis for RFS)</p> <p>Postmenopausal ER+ patients (n=179), multivariate Cox proportional hazard model analysis: Recurrence Rate (5y vs 2y), low ratio: 0.39 (CI 95% = 0.17–0.91), P value = 0.030 Test for interaction: P value = 0.035 Recurrence Rate (5y vs 2y)</p> <p>Postmenopausal ER+, node negative, patients (n=134), multivariate Cox proportional hazard model analysis: Recurrence Rate (5y vs 2y), low ratio: 0.27 (CI 95% = 0.10–0.72), P value = 0.0087</p> | <p>In this study the expression levels were normalized to b-actin using fresh frozen samples. Patients were collected from two distinct institutions; of 373 tumor samples analyzed, RNA expression data were obtained from 357 tumors</p> <p>The ratio or HOXB13 alone can predict the benefit of endocrine therapy, with a high ratio or a high expression rendering patients less likely to respond</p> |

BCP=breast cancer profiling; NR = not reported; RFS = relapse-free survival; OS = overall survival; ER= estrogen receptor; KM=Kaplan-Meier; CI= confidence interval.

Chapter 4. Discussion

Using the analytical framework described, we evaluated the evidence available on three commercially available gene expression based assays, and on the gene expression profiles underlying these tests. Specifically, our review focused on the MammaPrint[®] assay, based on the 70-gene prognostic signature developed by van't Veer and colleagues,^{21,25,58,59} on the Oncotype DX[™] assay, based on the 21-gene profile developed by Paik and colleagues,^{28,50,53} and on the Breast Cancer Profiling (BCP) assay, based on the two-gene ratio signature developed by Ma et al.^{61,64}

The first question, (is there any direct evidence that these tests in breast cancer patients lead to improvement in outcomes?) is defined as randomized clinical trials comparing the outcomes of patients following standard management to those of patients managed with the aid of the expression-based assays. No such studies have been conducted. Two prospective randomized trials are in progress: TAILORx³⁵ and MINDACT³⁶ were recently initiated to prospectively evaluate the clinical utility of Oncotype DX and MammaPrint, respectively. As described in Chapter 3, TAILORx will provide information on the appropriate RS threshold for recommending adjuvant chemotherapy, and will not directly assess the effect of clinical decisionmaking with and without the test. The data generated may allow indirect inferences to be made. MINDACT will allow more direct inferences on the clinical utility, since its will be compared directly to the use of a conventional risk index. For both trials, patient health outcomes will be endpoints.

The evidence available on the subsequent key questions allowed us to draw conclusions about the specific tests, as well as about the methodology of test development and current and future clinical uses of gene expression assays. Currently established methods for risk stratification of patients with breast cancer rely on a combination of prognostic factors like tumor size, grade, lymph node status, and presence of hormone receptors and the human growth factor receptor 2 (HER-2), such as the St. Gallen Consensus Guidelines⁵ or Adjuvant! Online.⁷ The latter also incorporates a nomogram to generate estimates of benefit from specific therapies. A critical question is how much gene expression-based tests add to standard risk assessment methods or guidelines. A second question is how clearly does the current evidence relate to the test's proposed use in a decisionmaking context, i.e., how well defined or homogeneous are the patient populations, in terms of their current therapy and decisions about future therapy? Is it clear how the test information should be implemented, i.e., using cutoffs, as a continuous score, or in combination with other indices? When viewed through the prism of clinical decisionmaking, the current evidence base for these technologies leaves many uncertainties.

Many aspects of expression-based predictors differ in qualitative ways from other kinds of risk predictors. First, the mechanism by which the expression of any particular gene, or combination thereof, is related to outcome is generally less well understood than with standard predictors, as are the methods by which the combinations are chosen. Gene expression levels are markers of activation or inactivation of complex biological processes. As Fan et al.⁷⁹ demonstrated, similar risk classifications can be achieved with predictors having few or no overlapping genes. Second, there is no "gold standard" for gene expression values; the technologies used here – RT-PCR and microarrays – represent the state of the art. In the end it is less analytic *validity* (i.e., proximity to a true value) but analytic *variability* (i.e., variation in the calculated value) that must be understood to predict whether investigational results are likely to

be similar to those produced in practice, and whether the results in practice are likely to be stable over time and with broader use. Third, we know little about the stability of the predictive value of such markers over populations with different genetic profiles. Arguments can be made that genetic predictors (particularly from tumors) are likely to be either more or less universal than physiologic ones, so there is still much to be learned about the generalizability of these rules. However, in spite of these differences, the latter half of the developmental pathway for these tests must follow the same principles and procedures as those for any multivariate clinical prediction rule. These have been outlined in detail in the clinical literature,^{94,95} enshrined in reporting guidelines,⁹⁶ and articulated with specific respect to expression-based predictors in a series of articles by Simon.^{68,71,97}

The three signatures and assays considered differ not only in the technologies used and their implementations, but also in the nature of the validation studies. An important distinction for all expression-based tests is that between the signature and the licensed test, as offered to a patient. Data about the actual tests offered to the patients are available only for MammaPrint and *Oncotype DX*, albeit more limited for the former. There is only one published study that used the two-gene index as it is implemented in its marketed version, the BCP assay,⁶¹ although it is not clear whether the lab performing the assay in this report was the same as the one with current rights to perform the test. The remaining reports considered the signature, with the expression of the two genes measured and combined in varying ways.^{61-63,72}

Recent publications have begun to address the analytic validity of the tests. There is now evidence about several aspects of gene expression measurements for two of the tests (MammaPrint and *Oncotype DX*).^{44,45,57,58} The public release of these data is useful as it supports the rationale behind two of the currently available assays and encourages development and publication of similar information for future assays. However, evidence about analytic features of the assays does not obviate the need for continuous monitoring of the experimental procedures involved with such testing. In this regard it is worth mentioning that the U.S. FDA Office of In Vitro Diagnostic evaluation and Safety (OIVD) is developing a Guidance Document on In Vitro Diagnostic Multivariate Index Assays (IVDMIAs) that will affect the development of future assays in the U.S. Moreover, the laboratories offering such assays, as any other laboratory providing diagnostic services must adhere to the Clinical Laboratory Improvement Act (CLIA).

Below follows the discussion of the specific tests and key questions considered in the present report, along with recommendations and conclusions.

Oncotype DX

Oncotype DX, the basis for the “recurrence score,” was first developed, then applied and used as an assay in investigational settings. All evidence about the RS (apart from the comparison study by Fan and colleagues⁷⁹ and the development studies^{44,48}) were obtained using the same assay that is offered to patients, with sample processing done in the same manner by the same laboratory.

Analytic Validity

Analytic validity evidence now exists for some of the operational/laboratory characteristics/procedures of this test, as well as about its reproducibility, although information about this latter point is limited to a few repeated analyses. These studies demonstrated that the

reproducibility of the test across different samples of the same block, and samples from different blocks, is reasonably high.⁴⁵ The test involves not only the simple assessment of the RNA levels by RT-PCR, but also the preparation of the RNA, following a central review of the specimens shipped to Genomic Health to check for tumor content. No direct evidence is available about the sample preparation aspect of the test, although there is indirect evidence from peer-reviewed literature in the form of the overall success rate of extracting analyzable mRNA, which appears to be fairly high. Centralization is a current strength of *Oncotype DX* with regard to reproducibility, but additional scrutiny may be needed if other laboratories offer such testing in the future.

Clinical Validity

“Clinical validity” is defined here as the ability of a prediction test to accurately predict risk. Whether or not those risk predictions differ enough to justify its use in a clinical setting (i.e., whether discrimination is sufficient) is a second issue. The clinical validity of *Oncotype DX* has been evaluated in various settings. The first validation study²⁸ used tamoxifen-treated women with ER positive, lymph node negative breast cancer, from the randomized clinical trial NSABP B-14. This study independently validated the prognostic value of the RS, which had been previously tested in the tamoxifen-treated population of NSABP B-20. Perhaps the most important aspect of this population is that it was clinically and prognostically well defined, in that everyone was presumed to be eligible for chemotherapy, and all subjects had similar treatment (i.e., tamoxifen), making for a relatively clear interpretation of the results in terms of both treatment biology and clinical decisionmaking. Predictors of response on a specified therapy are not necessarily prognostic factors independent of that therapy, so studies which mix treated and untreated patients, or patients differently treated, can produce results that do not apply well to either. While this study took place in the past, all measurements were done concurrently independently of the outcome, and so has evidential value quite close to that of a concurrent prospective study. The main issues raised by the non-concurrency are whether the 668 subjects examined were a representative sample of the more than 2000 in the original study, and the degree to which the findings in tamoxifen-treated women will apply to aromatase inhibitor treated patients today, the role of HER-2 testing and treatment, and whether there was anything clinically relevant about how the early stage cancer was diagnosed (e.g., clinically or by mammogram) that might differ today.

While this study reported hazard ratios for the RS in the presence of clinical predictors, it did not provide predictions by the RS cross-classified by those of standardized combination risk predictors to see exactly how many women would be re-classified in risk strata that might change decisions regarding chemotherapy. This information was however presented in poster form in 2004⁶⁵ and it showed that the RS had considerable predictive power beyond that of the St. Gallen or NCCN risk stratification guidelines (n.b. St. Gallen did not include HER-2 at that time). Another poster⁶⁶ showed the same information for Adjuvant! Online (Tables 13 and 14, Chapter 3). All of these cross-classifications showed that the greatest contribution of the test was likely to be in the reclassification of patients from high to low risk, i.e., in reducing the number of patients who might unnecessarily undergo chemotherapy. It also showed that optimal predictions probably would be achieved with a combination of both expression and clinical predictors. It must be stressed that the cross-classified risk of patients in the low-risk RS groups in these posters does not represent the lowest attainable risk; they are an artifact of the “low risk”

category threshold. Patients who had low absolute RS scores would be predicted to have lower risks than the “low risk” category average, and those lower risks are probably low enough for many women classified as high risk by other indices to change their treatment decisions. One very important remaining question is the degree to which the absolute observed risks in this population, particularly in these lowest risk groups, are similar to other populations, and to those whom it is currently being used, i.e. whether the calibration of the test predictions will vary. The low risk arm of the TAILORx trial, in which patients with a RS less than 10 will not be treated with adjuvant chemotherapy, will help address this question.

A second large study looked at the clinical performance of the *Oncotype* DX assay to predict breast cancer death (at 10 years) in a community-based population of ER positive, node-negative patients treated with tamoxifen, confirming the B14 results among the tamoxifen treated patients, and showing predictive value, albeit lesser, in ER positive patients not treated with tamoxifen.⁵⁰ The Esteva study,⁴⁸ which showed no predictive value of RS in a small population of patients who received neither tamoxifen nor chemotherapy, showed such anomalous results with standard predictors (i.e. higher grade predicting better prognosis) that its results cannot be regarded as reliable. Finally, the Fan study,⁷⁹ while testing the RS signature measured by microarrays and not the actual *Oncotype* DX test based on RT-PCR, showed good discriminatory power in a relatively large, independent dataset, albeit with a heterogeneous mix of treatments, receptor and nodal status. This is the same dataset on which the MammaPrint signature was developed.

These studies in combination provide fairly strong support for the clinical validity of the *Oncotype* DX test over and above standard predictors, in a well defined population (ER positive, lymph node negative, tamoxifen treated) with clear treatment indication (adjuvant chemotherapy). Exactly how much it adds, however, exactly what proportion of these patients would benefit from its use, and the stability of the observed risk in the various risk categories in other (or current) populations, is not as clear. Discussion will continue below about its use in a clinical setting.

Clinical Utility

Clinical utility is the degree to which a test is predictive of treatment benefit, and hence is a critical foundation for the use of a test in clinical decisionmaking. Prognostic ability itself speaks to this to some degree, as it puts a ceiling on the degree of clinical benefit. For example, if the 10 year distant relapse rate is 5 percent, by definition additional treatment cannot provide more than a 5 percent absolute benefit, and background knowledge about treatment efficacy tells us it will be less. So if the risk of distant recurrence can be reliably established as low enough, this has clinical utility in itself.

However, it is of considerable value to have a direct estimate of the degree of treatment benefit. This can only be done reliably in the context of RCTs, prospectively or retrospectively, as they assess treatment effect in an unbiased manner. This was addressed by Paik et al. in their study of the correlation of the RS with the degree of adjuvant chemotherapy benefit in the context of the NSABP-20 trial.⁵³ This showed that the chemotherapy benefit in ER positive, node negative patients randomized to tamoxifen versus tamoxifen plus chemotherapy was almost entirely restricted to those in the high risk RS category. The CIs in the low and intermediate risk categories were wide and included the possibility of benefit whereas the CI for the high risk group was narrow and showed clear benefit. A statistical interaction was also found with patient age, although those data were not reported. The only caveat is that the tamoxifen arm of this

population was part of the training set for the assay, although the outcome measure used in the training set was not treatment benefit. It is not clear whether the information in clinical predictors was optimally used (i.e., as continuous rather than dichotomous variables), but that is unlikely to have accounted for the degree of differential effect predicted by the RS. HER-2 positivity reportedly had no effect on the results. Several other studies evaluated the value of the RS information in different populations of patients to predict other correlates of treatment effect. For example, evaluation of pathologic response after preoperative chemotherapy^{49,55} supports clinical utility, although that was evaluated in patients in whom chemotherapy was already determined to be necessary.

The NSABP-20 study probably represents as strong evidence as can be derived from already existing data regarding the clinical utility of the *Oncotype* DX test. While prospective confirmation of these findings are definitely needed as well as analysis of existing patient samples from other completed trials, this provides reasonable justification in the interim for the use of the test by women in this specific population.

Use in clinical decisionmaking. One published study has reported the impact of using the RS on clinical management,⁵⁶ and there have been examinations of the economic implications of testing.⁷⁵ In general, studies showing that physicians change recommendations or that woman change treatment decisions in response to their *Oncotype* DX risk category are minimally informative if the study is not designed to specifically explore the woman's risk thresholds for making that decision. The reported study does not specify what information was conveyed to the patients, i.e., a risk score, the risk category, or the risk itself. If the latter, the number they were told is important to know. In the absence of this information, it is not possible to know the threshold of risk below which most women (or any given proportion) would forego chemotherapy, or conversely, the risks at which they would choose it. In the absence of such information, it cannot be known whether the study is effectively examining compliance with physician recommendations, careful weighing of risks and benefits, or the effect of test marketing.

There are still uncertainties about the optimal use of this test in practice. First, while the cut-offs are valuable for test validation purposes, it is not clear whether the current thresholds actually correspond to the cutoffs that would be derived using a formal decision-analytic approach based on utility assessments. For an individual woman, a risk based on her exact RS value would be preferable, since by definition, those with RS scores near the upper boundary of the "low risk" range have a predicted risk higher than the average of the group, and those with low scores have lower risk. The fact that the boundaries used in the studies may not be optimal for decisionmaking is seen in the different cut-offs used by the TAILORx trial, in which the low-risk group is defined as RS less-than or equal-to 10 instead of 18, and the high risk group is defined as greater than 25 instead of 30.

The second uncertainty is the optimal use of conventional predictors. While the RS has been shown to have more value than most predictors, the same studies show that clinical predictors retain predictive value, and clinical prediction models continue to evolve and improve. An improved prediction tool would involve a combination of the expression-based and clinical predictors, but this has not been systematically explored in any study, and absolute risks produced by regression models or stratified tables with all predictors included are generally not reported. As noted previously, cross-classification data using the most updated standardized clinical indices would be one form of such data, although those do not show the risk from combinations of the exact RS and clinical predictor score.

Cost-effectiveness. While our review highlights many gaps in what is known about the clinical utility of using gene expression profiling in women diagnosed with breast cancer, the review also revealed that little is known about the cost-effectiveness of using these tests. Once studies have demonstrated the clinical utility of these gene expression profiles, policy makers and health care providers will need information about the cost-effectiveness of those tests that have proven utility. Such information will be particularly important given the relatively high expected costs of the tests. *Oncotype DX*, for example, costs more than \$3000 for each use of the test.

In our review, we found three published studies that have addressed economic outcomes associated with use of the breast cancer gene expression tests. One study reported that using the 21-gene RT-PCR assay to reclassify patients would be cost-effective for those who were defined by 2005 NCCN criteria as low risk (\$31,452 per quality-adjusted life-year (QALY) gained) and would be cost-saving for those who were defined by NCCN criteria as high risk.⁶⁷ The EPC team had only moderate confidence in these projections because the study did not provide enough information about potential sources of bias in the analysis, allied with the fact that the study was supported by the manufacturer, which may introduce conflict of interest. The 2007 NCCN guideline now indicates that use of chemotherapy in these patients is optional, further diminishing the value of these projections.

The second study reported that use of the 21-gene RT-PCR assay was associated with a cost-utility ratio of \$4432 per QALY compared with use of tamoxifen alone, and a gain of 1.71 QALYs with net cost savings when compared with chemotherapy plus tamoxifen.⁷⁵ The EPC team had little confidence in this analysis, which was supported by the manufacturer, because it did not meet many of the standards that were used for appraising the quality of the analysis.

The third study compared the cost-effectiveness of the Netherlands Cancer Institute gene expression profiling (GEP) assay (MammaPrint) to the U.S. National Institutes of Health (NIH) guidelines for identification of early breast cancer patients who would benefit from adjuvant chemotherapy. The GEP assay was projected to yield a poorer quality-adjusted survival than the NIH guidelines (9.68 vs. 10.08 QALYs) and lower total costs (\$29,754 vs. \$32,636). To improve quality-adjusted survival, the GEP assay would need to have a sensitivity of at least 95 percent for detecting high risk patients while also having a specificity of at least 51 percent. The EPC team had confidence in the results of this analysis because it met most of the standards for appraising the quality of an economic analysis.

Since the overall body of evidence is inconclusive about the economic outcomes associated with use of breast cancer gene expression tests, this is an area that will require further investigation. Future economic analyses of validated tests should take into consideration existing guidelines for the performance and reporting of such analyses.⁴¹ Ideally, the analyses should be performed by investigators who have not received financial support from manufacturers of the tests.

Questions Regarding the Clinical Validity and Utility of the *Oncotype DX* Assay

1. Better information is needed about the predictions from combining the RS with current versions of standardized risk predictors, both in the form of cross-classification tables, and perhaps of regression-based combinations that optimize individual risk predictions. Formal development of cutoffs to optimize patient utility are also needed.
2. While *Oncotype DX* exhibits a fair bit of risk discrimination (i.e., separating patients into different risk groups), the stability across different populations of the observed absolute risk in patients with a given risk score (i.e., calibration) needs further study. Of greatest interest is the observed risk in the lowest risk groups, since the absolute level of this risk is critical for informed decisionmaking, and patients may forego chemotherapy on the basis of this information.
3. Data are currently available mainly for tamoxifen-treated patients and for those treated with cyclophosphamide-methotrexate-5-fluorocil chemotherapy. It is important to assess whether RS applies to other hormonal treatments such as aromatase inhibitors, as well as more contemporary chemotherapy regimens using taxanes and anthracyclines.
4. It is not clear whether RS can be used to help guide treatment of HER-2 positive patients and additional studies are needed, as most of these patients were classified in the high RS group in the initial trials.
5. While awaiting the TAILORx results, the findings of the Paik 2006⁵³ study predicting treatment benefit need independent confirmation, particularly for low and intermediate risk groups.
6. Studies examining the use of *Oncotype DX* should provide women and physicians with quantitative risk information and report how this alters clinical decisionmaking. The manner in which this risk information is presented should also be studied.

MammaPrint

Published evidence includes both reports about MammaPrint,⁵⁷⁻⁵⁹ as well as studies about the associated 70-gene signature. The manuscripts that used the signature provide useful information about the validity of the biological correlations underlying the profile and suggest that it can be used in a clinical setting, but cannot be considered to be a direct validation of the assay.

The assay is based on the gene signature first proposed in 2002²¹ by investigators at the Netherlands Cancer Institute, using 789 lymph node negative patients, younger than 55 years old, who did not carry a breast cancer gene (BRCA) mutation, and whose tumors were less than 5 cm in diameter. This signature was validated in a second study by the same group, using a series of 295 consecutive stage I or II breast cancer patients, who were either lymph node negative or positive, and who were younger than 53 years.²⁵ This validation was only partial, since the investigators included 61 of the 78 patients used to develop the prognosis profile. The MammaPrint test itself was further validated in a multicenter European study of 302 patients not treated with chemotherapy or tamoxifen, showing that it provided prognostic information beyond that of standard clinical-pathologic indices for those patients.⁵⁹ Recently, this signature was implemented as a commercial assay, and RNA available from the original cohorts were

reanalyzed, yielding consistent results.⁵⁸ It is the first prognostic test submitted to the FDA under its new, non-binding IVDMA guidance, and received approval in February 2007.

Analytic Validity

This assay is the first microarray-based test introduced in the field. Two recent papers addressed issues related to the reproducibility of the test within laboratories, as well as across laboratories. Such evidence however was obtained from a limited number of patients and using a moderate number of replication experiments. Results showed a good reproducibility within a laboratory, and a good degree of agreement across laboratories, although RNA labeling emerged as a possible source of variation capable of affecting the results. Whether this issue has an impact on risk classification was not thoroughly investigated, and thus the portability of the result of the assay from one laboratory to another still remains open. A second relevant point is the fact that the only validation study using the MammaPrint assay showed that only about 80 percent of specimens from the field (in this case 5 different European institutions) were analyzable, raising some concern about the analysis of fresh-frozen specimens. As more patients are analyzed by this test, the overall success rate may change. Finally, it must be noted that although this technology requires fresh rather than paraffin embedded specimens, Agendia performs a central pathologic review of the specimens as is performed with FFPE samples at Genomic Health, before evaluation with the test.

Clinical Validity

Overall, published evidence supports MammaPrint as a better predictor of the 5-year risk of distant recurrence than traditionally used tumor characteristics or algorithms. However, the cohorts in whom it was developed and validated are more clinically heterogeneous than those used for the *Oncotype DX* test, with a mix of lymph node status, ER status and current treatment. Additionally, evidence was derived only from patients younger than 55 to 60 years of age. Even so, it is interesting that it had 80 percent concordance with the array-based RS classification when applied to the same patients, although it remains to be seen how well it predicts in cohorts with the same degree of clinical and treatment homogeneity as used in the *Oncotype DX* development, and which differ from its training set. Evidence about its value in comparison with clinical predictors was assessed in a collaborative study among 5 different institutions in Europe, where data were compared to standard clinical predictors like Adjuvant!.⁵⁹ The area under the receiver operating characteristic curves was 0.68 for MammaPrint and 0.66 for the Adjuvant! Score. Such estimates indicate that both methods have apparently similar and modest discriminatory power in absolute terms. Similar results were obtained also using the ten year overall survival end point. However, when Adjuvant! and MammaPrint were cross-classified against each other, Adjuvant! had no additional predictive value. Adjustment for other predictors (St. Gallen and the Nottingham Prognostic Index) had a minimal effect on the regression coefficient of MammaPrint score or its significance, but no other data were reported on their incremental value. Of note, no significant heterogeneity in the hazard ratio estimates was shown among centers, although original hazard ratio estimates were significantly higher than those obtained in this validation study. The validation cohort had longer time of observation, included older women, and excluded patients who received adjuvant therapy.

Clinical Utility

No studies evaluated clinical utility of this test.

Use in clinical practice. No studies explored the use of this test in clinical practice.

In summary, MammaPrint is the first commercialized microarray-based gene expression profile with a prognostic purpose. The underlying signature has been evaluated in approximately 700 patients, although MammaPrint itself has only been evaluated in one study of 307 untreated patients. A reanalysis of the original training data of the signature using the marketed test showed a net reclassification of only one patient of 78.⁵⁸ It is unclear what population of patients would derive benefit from use of the test, and what the magnitude of that benefit would be. Prospective data from trials like MINDACT will be extremely valuable. Overall, published evidence supports MammaPrint as a better predictor of the risk of distant recurrence than traditionally used tumor characteristics or algorithms, but its performance in therapeutically homogeneous populations is not yet known with precision, and it is unclear for how many women the lowest predicted risks are low enough to forgo chemotherapy. No evidence is available to permit conclusions regarding the clinical utility of MammaPrint to select women who will benefit from chemotherapy.

To conclude, the literature on the 70-gene signature includes numerous studies that focused more on its biological underpinning and less on the clinical implications of this gene expression profile, although it has now received FDA approval for clinical use. It has been shown that this signature is maintained along the cancer progression process.⁸² This profile was directly investigated by two different platforms (microarray and RT-PCR,⁴⁸), and was successfully re-implemented in two distinct microarray platforms, showing that it has a fair degree of analytic robustness.

Here we summarize open questions as well as research gaps found in the evidence about the clinical validity and utility of the MammaPrint assay and the 70-gene signature.

1. The prognostic value of the 70-gene signature has been assessed in different populations facing different therapeutic choices. In the analysis by van de Vijver and colleagues, 130 of the 295 patients received adjuvant therapy in a non-randomized fashion. Patients in the original development cohort were not treated, and Buyse validated the marketed assay in untreated patients. It is not yet clear which are the optimal patient populations for the use of this test, exactly what its performance is in those populations, and how many of its predictions would result in different therapeutic decisions. Larger independent validation studies in therapeutically homogeneous groups would be very valuable.
2. Previous comments noted in the *Oncotype DX* summary apply here as well, including the presentation of data regarding the test in combination with standard predictors, the use of risk categories instead of a continuous risk measure, and the importance of confirming the stability of the test's calibration in different populations.

H/I Ratio Signature and Breast Cancer Profiling (BCP)

This test, licensed by AvicaraDX to Quest Diagnostic, Inc. is based on the two-gene ratio signature originally proposed by Ma and colleagues.⁶⁴ Specifically, the assay is based on the two-gene index that includes normalization to specific reference genes followed by a mathematical transformation.⁶¹ Overall, large collections of patients have been investigated using the signature, but its prognostic and predictive value has been inconsistent; strong in some studies, weak or absent in others. In the Fan study, in which the ratio based on the signature (not the marketed BCP test), it was completely non-predictive where both *Oncotype DX* and the *MammaPrint* signatures were. The reason for that may have been a technical failure of the array technology used to simulate the test,⁹⁸ or the test's value may be restricted to certain populations. The populations in which it has been developed have been heterogeneous, although stratified analyses were used. Differences have been found in its ability to predict in various subgroups of those populations, differences that are not consistent across studies. A major limitation of the evidence is that the signature has been formulated in a variety of ways, as a simple ratio, as an index, by normalizing to a different set of reference genes, or to a standard calibration RNA. In the 2006 study in which the index as is currently marketed was tested,⁶¹ statistical methods to find optimal cutoffs were applied, meaning that this assay still requires further external validation. We found no analytic validity data for the BCP assay.

In summary, while this test shows some promise, it must be regarded as being in a developmental phase. It was not clear in the Ma 2006⁶¹ study whether samples were processed by Quest Diagnostics, Inc. which holds the current license. There are a number of intriguing biological insights and plausible mechanisms to support the rationale for the test, but its consistent value in well-defined clinical settings has not yet been firmly established.

General Comments on Analytic Validity and Laboratory Quality Control

Until recently there were no multi-gene RNA-expression-based assay kits approved by the FDA for use in breast cancer. Such tests are currently offered as laboratory services (“home-brew test”) subject to CLIA general laboratory standards. In February 2007, and again in July 2007, the FDA published draft guidelines on regulation of IVDMIA, which cover tests combining complex algorithms and data from multiple laboratory tests. The release of these draft guidelines suggests that in the future these tests will be subject to FDA evaluation. Under this model, all the assays to be used to make medical decisions about therapeutic options will be regarded as Class II or III devices and will go through a Pre-Market Approval (PMA) process, and will require specific post-market revision. Based on such draft guidelines, *MammaPrint* receive IVDMIA approval upon their voluntary submission of data.^{38,39}

Nevertheless, analytic validity is an issue related to quality control in the laboratories where the test is carried out, and these data are not in the literature, but in the laboratories' log books. An effort has been made by Genomic Health, Inc. and Agendia to clarify the laboratory procedures and acknowledge critical issues, but periodic review and reporting of the procedures needs to be established to monitor the reproducibility of the procedures, success rates, and quality control indices.

A critical and often underappreciated analytic issue for the success of these tests is the way specimens are handled. Unlike DNA, RNA is unstable, so the length of time from excision to freezing or fixation, prolonged storage, and other factors related to specimen processing can lead to significant variability in the quality of mRNA available for expression profiling. Even if central labs offering the test are certified and use reliable procedures, preanalytic issues at the sending sites such as specimen acquisition and handling can potentially affect the results of the testing. Both the *Oncotype DX* and BCP use standard formalin fixed specimens, which tends to be stable, whereas MammaPrint requires fresh tissue. The use of fresh tissue required for gene array testing is challenging and, according to on-line information available from the Agendia website, careful procedures must be used when sampling the tumor to avoid necrotic parts and stromal tissue. Samples are reviewed centrally at Genomic Health and Agendia for tumor content, and BCP is performed after laser capture microdissection. Regardless of the technology used, standardized protocols, use of new reagents specifically designed to preserve mRNA for gene expression profiling, and reduction in RNA degradation (during sample processing, storage, and preparation) are important to assuring reliable measurements of mRNA levels for use in gene expression profiling.

Overall Implications and Recommendations

The discussion above covers issues specific to the tests under examination, but there are some larger issues whose consideration is motivated by this analysis that groups involved in assessing the value of these tests should be aware of.

Assay Validation

In general, it is clear that validation studies need to deal with populations for whom the decision-making implications of various risk groupings are clear. The studies examined herein have established the proof-of-concept that tumor gene expression has prognostic value, but for all tests except *Oncotype DX*, both validation and development studies have been on mixed populations, without sufficient sample sizes to stratify into large enough homogeneous groups to guide clinical decisionmaking. In addition, validation samples are often re-used by other investigators; the pool of such samples in the public domain needs to be greatly expanded.

Potential for Scale Problems

One problem that may be faced in the future is that of the consequences of an increase in demand for these tests. Scaling up the production could represent a challenge for the reproducibility and reliability of the tests in any setting, especially if more than one laboratory will offer the assays, since procedures to warrant inter-lab reproducibility will be needed. Not only analytical aspects will need monitoring, but also procedures involving specimen evaluation prior to testing. With a larger number of tests, for instance, the ability to reliably perform the central pathologic review might become an issue, while in the case of MammaPrint the availability of the current reference RNA could potentially become a limiting factor.

Genetic Variability and Gene Expression

It is unknown whether gene expression profiles are more or less likely than more traditional biomarkers to be generalizable beyond the populations in which they were initially developed. Gene expression may reflect fundamental biological tumor features, and thus be relatively stable across ethnic groups. However, gene expression patterns have also been associated with specific genetic mutations (i.e., BRCA1), indicating that specific DNA mutations or polymorphisms^{21,99} may affect the performance of a signature. This speaks to the importance of validating these tests in populations with varying genetic background. Biological and genetic evidence potentially addressing these issues is expected to become available in the form of single nucleotide polymorphism (SNP) arrays coupled to expression arrays.

The Need for Databases, Reproducibility, and Standards

MammaPrint® is the first assay based on microarrays that has completed the path from the bench to FDA approval for clinical application. For data storage, the MIAME standards³² represent the basis for the proper collection and storage of microarray data, and should be used to develop procedures going forward for the archiving of the tests performed in real patients, much as databases have been developed to facilitate outcomes research to complement clinical trials. Consideration should be given to the development of databases with complete data on each patient (absent identifiers), including all the analyses performed, laboratory logs, the raw and processed data, and all the information about procedures and analyses that have been performed to produce a risk estimate from a tumor sample. These apply equally to the other two assays, differing only in the type of data that would be stored.

Where Is the Field Going?

The current evidence for the feasibility of such gene expression based tests in clinical settings, along with the demand for better tools to manage patients, is leading to both an evolution of the available tests, and the addition of novel alternative tests. The number of publications is growing, and several alternative signatures not considered here have already been proposed for breast cancer as well as for other neoplasms. We can expect many new tests, as well as new uses for the assays that already exist. More genes might be added to the signatures, and in the particular case of MammaPrint this will be possible without changing the experimental procedures, since the array contains thousands more genes than the ones that are incorporated in the 70-gene signature. In this regard, we might also expect other modifications: subsets of the current signatures might be proposed as alternatives to current clinical risk factors, or be proposed in different populations or for different purposes. For *Oncotype DX*, a natural evolution could be related to its use as an alternative to immunohistochemistry and/or pathology to evaluate tumor Grade, S-phase index, ER, PR, and HER-2 expression, since such genes are part of the set included in the assay. Reporting of individual gene expression results may also prove useful. A great deal more work needs to be done on the prediction of therapeutic benefit, which is the ultimate goal of all such tests.

“Comparative Effectiveness” Studies

The emphasis in virtually all of the papers and in our evidence assessment is on the establishment of the value of each of these predictors over standard clinical predictors. However, as gene expression tests mature and proliferate, an important question will be how they compare to each other, and whether there is value in their combination. In the therapeutic domain, this has been called “comparative effectiveness” research. Such research has traditionally been difficult to fund by government or by industry, because it may not hold out as much therapeutic promise as new discoveries, and because industry understandably is not anxious to fund head-to-head comparisons with competitive products. This same dynamic could easily take hold in the risk prediction arena, with a proliferation of licensed prediction indices without any clear notion of what new ones are contributing over previous tests. Development of future expression-based predictors should make clear their incremental value over pre-existing methods. In the absence of better oversight of test development, physicians and patient are likely to be awash in new tests that all claim to offer similar guidance, or perhaps new guidance in previously neglected clinical subsets, with no way to sort out those claims.

Conclusion

The introduction of these gene-expression tests have ushered in a new era in which many conventional clinical markers and predictors may be seen merely as surrogates for more fundamental genetic and physiologic processes. The multidimensional nature of these predictors demands both large numbers of clinically homogeneous patients to be used in the validation process, and exceptional rigor and discipline. Every study provides an opportunity to tweak a genetic signature, but we must find the right balance between speed of innovation and development of scientifically and clinically reliable tools. Going forward, it will be important to harness, if possible, as much genetic and clinical information on patients who undergo these tests to facilitate each goal without unduly sacrificing the other.

References and Included Studies

1. Jemal A, Siegel R, Ward E et al. Cancer statistics, 2007. *CA Cancer J Clin* 2007;57(1):43-66.
2. Berry DA, Cirincione C, Henderson IC et al. Estrogen-receptor status and outcomes of modern chemotherapy for patients with node-positive breast cancer. *JAMA* 2006;295(14):1658-67.
3. Eifel P, Axelson JA, Costa J et al. National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, November 1-3, 2000. *J Natl Cancer Inst* 2001;93(13):979-89.
4. National Institutes of Health (NIH) Consensus Development Criteria web site. Available at: <http://consensus.nih.gov/2000/2000AdjuvantTherapyBreastCancer114html.htm>. Accessed July 25, 2007.
5. Goldhirsch A, Glick JH, Gelber RD et al. Meeting Highlights: International Expert Consensus on the Primary Therapy of Early Breast Cancer 2005. *Breast* 2005;14(6):643.
6. Carlson RW, Anderson BO, Burstein HJ et al. Invasive breast cancer. *J Natl Compr Canc Netw* 2007;5(3):246-312.
7. Ravdin PM, Siminoff LA, Davis GJ et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol* 2001;19(4):980-91.
8. Adjuvant!, Inc. Adjuvant! Online. Available at: <http://www.adjuvantonline.com>. Accessed July 25, 2007
9. Wolff AC, Hammond ME, Schwartz JN et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *J Clin Oncol* 2007;25(1):118-45.
10. Braxton S and Bedilion T. The integration of microarray information in the drug development process. *Curr Opin Biotechnol* 1998;9(6):643-9.
11. Mirnics K. Microarrays in brain research: the good, the bad and the ugly. *Nat Rev Neurosci* 2001;2(6):444-7.
12. Mirnics K, Middleton FA, Lewis DA et al. Analysis of complex brain disorders with gene expression microarrays: schizophrenia as a disease of the synapse. *Trends Neurosci* 2001;24(8):479-86.
13. Schulze A and Downward J. Navigating gene expression using microarrays--a technology review. *Nat Cell Biol* 2001;3(8):E190-5.
14. van Berkum NL and Holstege FC. DNA microarrays: raising the profile. *Curr Opin Biotechnol* 2001;12(1):48-52.
15. DeRisi J, Penland L, Brown PO et al. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet* 1996;14(4):457-60.
16. Alizadeh A, Eisen M, Davis RE et al. The lymphochip: a specialized cDNA microarray for the genomic-scale analysis of gene expression in normal and malignant lymphocytes. *Cold Spring Harb Symp Quant Biol* 1999;64:71-8.
17. Alizadeh AA, Ross DT, Perou CM et al. Towards a novel classification of human malignancies based on gene expression patterns. *J Pathol* 2001;195(1):41-52.
18. Rew DA. DNA microarray technology in cancer research. *Eur J Surg Oncol* 2001;27(5):504-8.
19. Hu Z, Fan C, Oh DS et al. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* 2006;7:96.
20. Sorlie T, Perou CM, Tibshirani R et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;98(19):10869-74.
21. van 't Veer LJ, Dai H, van de Vijver MJ et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415(6871):530-6.
22. Chang HY, Nuyten DS, Sneddon JB et al. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci U S A* 2005;102(10):3738-43.
23. Sotiriou C, Wirapati P, Loi S et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 2006;98(4):262-72.

24. Huang E, Cheng SH, Dressman H et al. Gene expression predictors of breast cancer outcomes. *Lancet* 2003;361(9369):1590-6.
25. van de Vijver MJ, He YD, van't Veer LJ et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;347(25):1999-2009.
26. Bustin SA. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* 2000;25(2):169-93.
27. Schena M, Shalon D, Davis RW et al. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995; 270(5235):467-70.
28. Paik S, Shak S, Tang G et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;351(27):2817-26.
29. Baldwin D, Crane V, and Rice D . A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants. *Curr Opin Plant Biol* 1999;2(2):96-103.
30. Watson A, Mazumder A, Stewart M et al. Technology for microarray analysis of gene expression. *Curr Opin Biotechnol* 1998;9(6):609-14.
31. Schena M, Heller RA, Theriault TP et al. Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol* 1998;16(7):301-6.
32. Brazma A, Hingamp P, Quackenbush J et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001;29(4):365-71.
33. Bammler T, Beyer RP, Bhattacharya S et al. Standardizing global gene expression analysis between laboratories and across platforms. *Nat Methods* 2005;2(5):351-6.
34. Irizarry RA, Warren D, Spencer F et al. Multiple-laboratory comparison of microarray platforms. *Nat Methods* 2005;2 (5):345-50.
35. National Cancer Institute. The TAILORx Breast Cancer Trial. Available at: <http://www.cancer.gov/clinicaltrials/digestpage/TAILORx>. Accessed August 14, 2007.
36. TransBIG. MINDACT. Available at: <http://www.breastinternationalgroup.org/TransBIG/Mindact.aspx>. Accessed August 15, 2007.
37. Food and Drug Administration, Center for Devices and Radiological Health. Available at: <http://www.fda.gov/cdrh/>. Accessed July 25, 2007
38. Food and Drug Administration, Center for Devices and Radiological Health. 510(k) Substantial Equivalence Determination Decision Summary, No. k062694. Available at: <http://www.fda.gov/cdrh/reviews/K062694.pdf>. Accessed July 25, 2007.
39. Food and Drug Administration, Center for Devices and Radiological Health. 510(k) Submission for MammaPring Service in the U.S. Summary, No. k070675. Available at: <http://www.fda.gov/cdrh/pdf7/K070675.pdf>. Accessed July 25, 2007.
40. Berlin JA. Does blinding of readers affect the results of meta-analyses? University of Pennsylvania Meta-analysis Blinding Study Group. *Lancet* 1997;350(9072):185-6.
41. Philips Z, Ginnelly L, Sculpher M et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess* 2004;8(36):iii-iv, ix-xi, 1-158 .
42. Bossuyt PM, Reitsma JB, Bruns DE et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Fam Pract* 2004;21(1):4-10.
43. McShane LM, Altman DG, Sauerbrei W et al. REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer* 2005;93(4):387-91.
44. Cronin M, Pho M, Dutta D et al. Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. *Am J Pathol* 2004;164(1):35-42.
45. Cronin M, Sangli C, Liu ML et al. Analytical Validation of the Oncotype DX Genomic Diagnostic Test for Recurrence Prognosis and Therapeutic Response Prediction in Node-Negative, Estrogen Receptor-Positive Breast Cancer. *Clin Chem* 2007;53(6):1084-91.
46. Braxton S and Bedilion T. The integration of microarray information in the drug development process. *Curr Opin Biotechnol* 1998;9(6):643-9.

47. Cobleigh MA, Tabesh B, Bitterman P et al. Tumor gene expression and prognosis in breast cancer patients with 10 or more positive lymph nodes. *Clin Cancer Res* 2005;11(24 Pt 1):8623-31.
48. Esteva FJ, Sahin AA, Cristofanilli M et al. Prognostic role of a multigene reverse transcriptase-PCR assay in patients with node-negative breast cancer not receiving adjuvant systemic therapy. *Clin Cancer Res* 2005;11(9):3315-9.
49. Gianni L, Zambetti M, Clark K et al. Gene expression profiles in paraffin-embedded core biopsy tissue predict response to chemotherapy in women with locally advanced breast cancer. *J Clin Oncol* 2005;23(29):7265-77.
50. Habel LA, Shak S, Jacobs MK et al. A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. *Breast Cancer Res* 2006;8(3):R25.
51. Mina L, Soule SE, Badve S et al. Predicting response to primary chemotherapy: gene expression profiling of paraffin-embedded core biopsy tissue. *Breast Cancer Res Treat* 2006.
52. Bertone P, Stolc V, Royce TE et al. Global identification of human transcribed sequences with genome tiling arrays. *Science* 2004;306(5705):2242-6.
53. Paik S, Tang G, Shak S et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* 2006;24(23):3726-34.
54. Sorlie T, Tibshirani R, Parker J et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003;100(14):8418-23.
55. Chang JC, Makris A, Gutierrez MC et al. Gene expression patterns in formalin-fixed, paraffin-embedded core biopsies predict docetaxel chemosensitivity in breast cancer patients. *Breast Cancer Res Treat* 2007.
56. Oratz R and Dev P. Impact of Oncotype DX™ Recurrence Score on Decision Making in Early-Stage Breast Cancer. *Journal of Oncology Practice* in press.
57. Ach RA, Floore A, Curry B et al. Robust interlaboratory reproducibility of a gene expression signature measurement consistent with the needs of a new generation of diagnostic tools. *BMC Genomics* 2007;8(1):148.
58. Glas AM, Floore A, Delahaye LJ et al. Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics* 2006;7:278.
59. Buyse M, Loi S, van't Veer L et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* 2006;98(17): 1183-92.
60. Hughes TR, Marton MJ, Jones AR et al. Functional discovery via a compendium of expression profiles. *Cell* 2000;102(1):109-26.
61. Ma XJ, Hilsenbeck SG, Wang W et al. The HOXB13:IL17BR expression index is a prognostic factor in early-stage breast cancer. *J Clin Oncol* 2006;24(28):4611-9.
62. Goetz MP, Suman VJ, Ingle JN et al. A two-gene expression ratio of homeobox 13 and interleukin-17B receptor for prediction of recurrence and survival in women receiving adjuvant tamoxifen. *Clin Cancer Res* 2006;12(7 Pt 1):2080-7.
63. Jerevall PL, Brommesson S, Strand C et al. Exploring the two-gene ratio in breast cancer-independent roles for HOXB13 and IL17BR in prediction of clinical outcome. *Breast Cancer Res Treat* 2007.
64. Ma X-J, Wang Z, Ryan PD et al. A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* 2004;5(6):607-616.
65. Paik S, Shak S, and Tang G. Risk classification of breast cancer patients by the recurrence score assay: comparison to guidelines based on patient age, tumor size, and tumor grade. Abstract presented at Annual San Antonio Breast Cancer Symposium; December 8-11, 2004. San Antonio, TX. Abstract 104.
66. Bryant, et al., 2005. Toward a More Rational Selection of Tailored Adjuvant Therapy. Poster Presentation. St. Gallens Conference, March 2005.
67. Hornberger J, Cosler LE, and Lyman GH. Economic analysis of targeting chemotherapy using a 21-gene RT-PCR assay in lymph-node-negative, estrogen-receptor-positive, early-stage breast cancer. *Am J Manag Care* 2005; 11(5):313-24.
68. Simon Z, Sipka S, Gergely L et al. Investigation of monoclonal gammopathy of undetermined significance: a single-centre study. *Clin Lab Haematol* 2006;28(3):164-9.

69. Reid JF, Lusa L, De Cecco L et al. Limits of predictive models using microarray data for breast cancer clinical treatment outcome. *J Natl Cancer Inst* 2005;97(12):927-30.
70. Sotiriou C, Neo SY, McShane LM et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A* 2003;100(18):10393-8.
71. Simon R. Development and validation of therapeutically relevant multi-gene biomarker classifiers. *J Natl Cancer Inst* 2005;97(12):866-7.
72. Jansen MP, Sieuwerts AM, Look MP et al. HOXB13-to-IL17BR expression ratio is related with tumor aggressiveness and response to tamoxifen of recurrent breast cancer: a retrospective study. *J Clin Oncol* 2007;25(6):662-8.
73. Sieuwerts AM, Meijer-van Gelder ME, Timmermans M et al. How ADAM-9 and ADAM-11 differentially from estrogen receptor predict response to tamoxifen treatment in patients with recurrent breast cancer: a retrospective study. *Clin Cancer Res* 2005;11(20):7311-21.
74. Carlson RW, Anderson BO, Burstein HJ et al. Breast cancer. *J Natl Compr Canc Netw* 2005;3(3):238-89.
75. Lyman GH, Cosler LE, Kuderer NM et al. Impact of a 21-gene RT-PCR assay on treatment decisions in early-stage breast cancer: an economic analysis based on prognostic and predictive validation studies. *Cancer* 2007; 109(6):1011-8.
76. Oestreicher N, Ramsey SD, Linden HM et al. Gene expression profiling and breast cancer care: what are the potential benefits and policy implications? *Genet Med* 2005;7(6):380-9.
77. Brauer CA, Rosen AB, Greenberg D, Neumann PJ. Trends in the measurement of health utilities in published cost-utility analyses. *Value Health* 2006; 9(4):213-8.
78. Torrance GW, Feeny D. Utilities and quality-adjusted life years. *Int J Technol Assess Health Care* 1989; 5(4):559-75.
79. Fan C, Oh DS, Wessels L et al. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 2006;355(6):560-9.
80. Espinosa E, Vara JA, Redondo A et al. Breast cancer prognosis determined by gene expression profiling: a quantitative reverse transcriptase polymerase chain reaction study. *J Clin Oncol* 2005;23(29):7278-85.
81. Eden P, Ritz C, Rose C et al. "Good Old" clinical markers have similar power in breast cancer prognosis as microarray gene expression profilers. *Eur J Cancer* 2004;40(12):1837-41.
82. Weigelt B, Hu Z, He X et al. Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. *Cancer Res* 2005;65(20):9155-8.
83. Nuyten DS, Kreike B, Hart AA et al. Predicting a local recurrence after breast-conserving therapy by gene expression profiling. *Breast Cancer Res* 2006;8(5):R62.
84. Chang HY, Sneddon JB, Alizadeh AA et al. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol* 2004;2(2):E7.
85. Chi J-T, Wang Z, Nuyten DSA et al. Gene expression programs in response to hypoxia: Cell type specificity and prognostic significance in human cancers. *PLoS Med.* 2006;3(3):395-409.
86. Naderi A, Teschendorff AE, Barbosa-Morais NL et al. A gene-expression signature to predict survival in breast cancer across independent data sets. *Oncogene* 2006.
87. Sun Y, Goodison S, Li J et al. Improved breast cancer prognosis through the combination of clinical and genetic markers. *Bioinformatics* 2007;23(1):30-7.
88. Esteva FJ, Sahin AA, Rassidakis GZ et al. Jun Activation Domain Binding Protein 1 Expression Is Associated with Low p27Kip1 Levels in Node-Negative Breast Cancer. *Clin. Cancer Res.* 2003;9(15):5652-5659.
89. Wang J, Buchholz TA, Middleton LP et al. Assessment of histologic features and expression of biomarkers in predicting pathologic response to anthracycline-based neoadjuvant chemotherapy in patients with breast carcinoma. *Cancer* 2002 ;94(12):3107-14.
90. Harvey JM, Clark GM, Osborne CK et al. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. *J Clin Oncol* 1999;17(5):1474-81.
91. Mohsin SK, Weiss H, Havighurst T et al. Progesterone receptor by immunohistochemistry and clinical outcome in breast cancer: a validation study. *Mod Pathol* 2004;17(12):1545-54.

92. Allred DC, Harvey JM, Berardo M, Clark GM. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol* 1998; 11(2):155-68.
93. Allred DC, Harvey JM, Berardo M et al. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol* 1998;11(2):155-68.
94. Laupacis A, Sekar N, and Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA* 1997;277(6):488-94.
95. Harrell FE Jr, Lee KL, and Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15(4):361-87.
96. McShane LM, Altman DG, Sauerbrei W et al. Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst* 2005;97(16):1180-4.
97. Dupuy A and Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Natl Cancer Inst* 2007;99(2):147-57.
98. Goetz MP, Ingle JN, and Couch FJ. Gene-expression-based predictors for breast cancer. *N Engl J Med* 2007;356(7):752; author reply 752-3.
99. Hedenfalk I, Duggan D, Chen Y et al. Gene-expression profiles in hereditary breast cancer. *N Engl J Med* 2001;344(8):539-48.
100. Auer H, Lyianarachchi S, Newsom D et al. Chipping away at the chip bias: RNA degradation in microarray analysis. *Nat Genet* 2003;35(4):292-3.
101. Baldwin D, Crane V, and Rice D. A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants. *Curr Opin Plant Biol* 1999;2(2):96-103.
102. Klein JP. Small sample moments of the estimators of the variance of the Kaplan-Meier and Nelson-Aalen estimators. *Scand J Stat* 1991;18:333-40

Appendixes

Appendix A: List of Acronyms

| Acronym | Definition |
|---------|---|
| AHRQ | Agency for Healthcare Research and Quality |
| ANOVA | Analysis of variance |
| AUC | Area under the receiver-operating-characteristic curve |
| BCP | Breast Cancer Profiling |
| BCT | Breast conserving therapy |
| BIG | Breast International Group |
| BRCA | Breast cancer gene |
| CCD | Charge Coupled Devices |
| CDC | Centers for Disease Control and Prevention |
| cDNA | Complementary DNA |
| CENTRAL | The Cochrane Central Register of Controlled Trials |
| CI | Confidence Interval |
| CLIA | Clinical Laboratory Improvement Act |
| CMTP | Center for Medical Technology Policy |
| CR | Complete response |
| CT | Cycle threshold |
| CV | Coefficient of variation |
| DFS | Disease free survival |
| DNA | Deoxyribonucleic acid |
| DRFS | Distant recurrence-free survival |
| EGAPP | Evaluation of Genomic Applications in Practice and Prevention |
| EPC | The Evidence-based Practice Center |
| ER | Estrogen receptor |
| FDA | Food and Drug Administration |
| FFPE | Formalyn-fixed paraffin-embedded |
| FISH | Fluorescent <i>in situ</i> hybridization |
| FRET | Förster Resonance Energy Transfer |
| GEP | Gene expression profiling |
| HER2 | Human epidermal growth factor receptor 2 |
| HR | Hormone Receptors |
| IHC | Immunohistochemical |
| INT | Italian National Cancer Institute of Milan, Italy |
| IVDMIAs | In Vitro Diagnostic Multivariate Index Assays |
| JHU | Johns Hopkins University |
| LCM | Laser-capture micro dissected |
| LMC | Laser micro-dissection |
| LOD | Limit of detection |
| LOQ | Limit of quality |
| MeSH | Medical subject heading |
| MIAME | Minimum Information About a Microarray Experiment |
| MINDACT | Microarray for Node-Negative Disease may Avoid Chemotherapy |
| mRNA | Messenger ribonucleic acid |

| Acronym | Definition |
|----------------|--|
| NCCN | National Comprehensive Cancer Network |
| NCCTG | North Central Cancer Treatment Group |
| NCI | National Cancer Institute |
| NHG | Nottingham Histologic Grade |
| NIH | National Institutes of Health |
| NPI | Nottingham Prognostic Index |
| NSABP | National Surgical Adjuvant Breast and Bowel Project |
| OR | Odds Ratio |
| OS | Overall survival |
| OVID | Office of In Vitro Diagnostic evaluation and Safety |
| pCR | Complete pathological response |
| PCR | Polymerase chain reaction |
| PD | Progressive disease |
| PFS | Progression free survival |
| PMA | Pre-Market Approval |
| PR | Progesterone receptor |
| PR | Partial response |
| QALY | Quality-adjusted life-year |
| RCT | Randomized Controlled Trial |
| RECIST | Response Evaluation Criteria in Solid Tumors |
| REMARK | Reporting recommendations for tumour MARKer prognostic studies |
| RFS | Relapse Free Survival |
| RNA | Ribonucleic acid |
| ROC | Receiver operating characteristic |
| RR | Relative risk |
| RS | Recurrence Score |
| RT-PCR | reverse transcriptase polymerase chain reaction |
| SD | Standard deviation |
| SNP | Single nucleotide polymorphisms |
| STARD | Standards for Reporting of Diagnostic Accuracy |
| TAILORx | Trial Assigning Individualized Options for Treatment |
| TBCI | The North American Breast Cancer Intergroup |
| TRANSBIG | Translating molecular knowledge into early breast cancer management: building on the BIG (Breast International Group) network for improved treatment tailoring |
| TTM | Time to distant metastases |
| VEGF | Vascular epithelial growth factor |

Appendix B: Glossary

Cycle Threshold (Ct, CT, C^t)

In an RT-PCR reaction template, the relative ratios of products and reagents vary. At the beginning of the process, reagents are in excess, and template and products are at low concentrations and do not compete with primer binding, so that the amplification proceeds at a constant, exponential rate. After this initial phase, the process enters a linear phase of amplification, due to competition of product renaturation with primer binding. In late reaction cycles, the amplification reaches a plateau phase and no more products accumulate. To achieve accuracy and precision, it is necessary to collect quantitative data during the exponential phase of amplification, since in this phase amplification is extremely reproducible. In RT-PCR, this process is automated and measurements are made at each cycle. The ‘cycle threshold’ is the cycle of the RT-PCR reaction corresponding to the beginning of the exponential phase of amplification.

DNA Microarray

A DNA microarray (also commonly referred to as “gene chip,” “DNA chip”) is a collection of microscopic DNA spots (defined “features”), commonly representing single genes or transcripts, arrayed on a solid surface by covalent attachment to chemically suitable matrices, or directly synthesized on them. DNA microarrays use DNA as part of their detection system. Qualitative or quantitative measurements with DNA microarrays use the selective nature of DNA-DNA or DNA-RNA hybridization under high-stringency conditions and fluorophore-based detection. DNA arrays are commonly used for gene expression profiling, i.e., monitoring expression levels of thousands of genes simultaneously, or for comparative genomic hybridization.

Gene Annotation

Gene annotation is the body of information that is associated with genes, as well as the process involved with the generation and maintenance of such information. Molecular biology and bioinformatics have faced the need for DNA annotation since the 1980s. Today a number of genomic and proteomic annotation projects have made this information publicly available.

Gene Expression

Gene expression refers to the translation of the information encoded in a gene into an RNA transcript. Expressed transcripts include messenger RNAs (mRNA) translated into proteins, as well as other types of RNA, such as transfer RNA (tRNA), ribosomal RNA (rRNA), micro RNA (miRNA), and non-coding RNA (ncRNA), that are not translated into protein. Gene expression is a highly specific process by which cells switch genes on and off in a timely manner, according to their state. The study of mRNA expression in a cell is an indirect way to study the proteins counterpart.

Gene Expression Classifier

The term classifier is derived from the field of machine learning. The goal of classification is to group items that have similar feature values into groups. Usually, in the context of gene expression analysis, a classifier is a composite algorithm that achieves patients classification by using gene expression measurements.

Gene Expression Profiling

This term refers to any genomic techniques that measure the fraction of the genes that is expressed in a specific sample. This definition refers to techniques that allow the assessment of more than one gene at a time, especially microarray and real time RT-PCR.

Gene expression profile: This is any set of genes for which the expression in a specific sample is known. A gene expression profile may account for a variable number of genes, and the corresponding expression values may be obtained by different techniques. Gene expression profiles can be associated, by various techniques, to phenotypes.

Gene expression pattern: This is an equivalent term currently in use to refer to “gene expression profile.”

Gene expression signature: This is an equivalent term currently in use to refer to a specific “gene expression profile,” usually associated with a specific phenotype.

Genome

In biology the genome of an organism is its whole hereditary information and is encoded in the DNA (for some viruses, RNA). This includes both the genes encoding for proteins, as well as the non-coding sequences of the DNA. The term, coined in 1920 by Hans Winkler, is the fusion of the words gene and chromosome. The study of the global properties of genomes is usually referred to as ‘genomics’, which distinguishes it from genetics, which generally studies the properties of single genes or groups of genes.

Laser Capture Microdissection

Laser Capture Microdissection (LCM) is a method for isolating pure cells of interest from specific regions of tissue sections. In this procedure a special film is applied on tissue sections that are analyzed under the microscope. When the cells of choice are identified, the operator can use a laser to dissect the cells and transfer them off of the film leaving all unwanted cells behind in the tissue section. LCM does not alter or damage the morphology and chemistry of the sample collected from which is possible to prepare DNA, RNA and/or protein. LCM can be performed on a variety of tissue samples, including blood smears, cytologic preparations, cell cultures and frozen and paraffin embedded archival tissue.

MIAME

MIAME (Minimum Information About a Microarray Experiment) is a standard for reporting microarray experiments. It is intended to specify all the information necessary to interpret the results of the experiment unambiguously and to reproduce the experiment. While the standard defines the content desired for reports, it does not specify the format in which this data should be presented. There are a number of file formats for representing this data, and both public and subscription-based repositories for such experiments.

Normalization

In an experimental context, normalizations are used to standardize data to enable differentiation between real (biological) variations and variations due to the measurement process. In gene expression analysis (by DNA microarray or RT-PCR), normalization refers to the process of identifying and removing the systematic effects, bringing the data from different samples onto a

common scale. Several alternative methods and approaches to perform normalization exist both for RT-PCR and DNA microarray.

Oligonucleotide

Oligonucleotides are short sequences of nucleotides (RNA or DNA), typically with twenty or fewer bases, although automated synthesizers allow the synthesis of oligonucleotides up to 200 bases. The length of a synthesized base is usually denoted by the suffix 'mer': for example, a fragment of 25 bases would be called a 25-mer. Oligonucleotides are used as probes to detect complementary DNA or RNA molecules. Specific DNA oligonucleotides are used in the PCR, and in this instance, they are referred to as “primers,” since they generate a place for the DNA polymerase to bind and extend the primers themselves, by the addition of nucleotides to make a copy of the target sequence. Oligonucleotides are may be referred to as “oligos.”

Platform

In the context of gene expression profiling analysis the term “platform” is often used to refer to the technology, instruments, and protocols used to measure gene expression. In this sense real time RT-PCR, cDNA microarrays, and oligonucleotide microarrays represent different platforms.

Polymerase Chain Reaction (PCR)

PCR is a molecular biology technique for isolating and exponentially amplifying a DNA sequence of interest in vitro via enzymatic replication. This technique has been extensively modified to perform a wide array of tasks, and it is now a common tool used in medical and biological research. PCR is now used to obtain the sequence of genes, to diagnose hereditary diseases, identify genetic fingerprints (forensics medicine), detect infectious diseases, and create transgenic organisms. Coupled to “reverse transcription” it is used to amplify RNA molecules.

Primer

A primer is a nucleic acid strand or a related molecule that serves as a starting point for DNA replication. A primer is required because most DNA polymerases cannot begin synthesizing a new DNA strand from scratch, but can only add to an existing strand of nucleotides. In most natural DNA replication, the ultimate primer for DNA synthesis is a short strand of RNA. This RNA is produced by “primase,” and is later removed and replaced with DNA by a DNA polymerase. Many laboratory techniques of biochemistry and molecular biology that involve DNA polymerases, such as DNA sequencing and polymerase chain reaction, require primers. The primers used for these techniques are usually short, chemically synthesized DNA molecules with a length about twenty bases.

Probe

In molecular biology, a hybridization probe is a fragment of DNA of variable length, which is used to detect the presence of nucleotide sequences that are complementary to the sequence in the probe. The complementary sequences are referred to as “targets.” The hybridization probe is usually labeled radioactively, or with immunological or fluorescent markers. The labeled probe is then denatured (by heating) into single DNA strands and hybridized to target DNA (Southern blotting) or RNA (Northern blotting) immobilized on a membrane or in situ. In a DNA microarray the hybridization scheme is reversed and the probes are attached to a solid surface, while the labeled targets are in the reaction solution. Similarly, in real time RT-PCR, probes are fragments of DNA that fluoresce when hybridized to the complementary investigated RNA molecule.

Proteome

The term proteome was coined by Mark Wilkins in 1994, as the fusion between proteins and genome. This term refers to the entire set of proteins expressed by a genome, cell, tissue or organism at a given time under defined conditions. The proteome is larger and more complex than the genome, especially in eukaryotes, in the sense that there are more proteins than genes. This is due to alternative splicing of genes and post-translational modifications like glycosylation or phosphorylation.

Real Time Reverse Transcriptase Polymerase Chain Reaction (RT-PCR)

Real-time RT-PCR is a molecular biology technique that allows the amplification and the quantification in real time of defined RNA molecules from specific specimens. This technology has been used for several years in research and clinical settings to measure RNA molecules. In the first step DNA, copies of the investigated RNA molecules present in the template are obtained by a reaction named reverse transcription. Then DNA amplification is obtained using PCR, while the quantification of the accumulating DNA product is accomplished by the use of specific fluorescent reagents. The quantification of the target RNA molecule is based on the analysis of the accumulation curve of the complementary DNA, as measured by the fluorescence detected at each cycle of the reaction.

Reverse Transcription

In biochemistry, reverse transcription is the enzymatic reaction induced on by the RNA-dependent DNA polymerase. This enzyme, also known as reverse transcriptase, is a DNA polymerase enzyme that copies single-stranded RNA into DNA. This process is the reverse of normal transcription, which involves the synthesis of RNA from DNA.

Ribonuclease

This type of enzyme, abbreviated commonly as RNase, is a nuclease that catalyzes the hydrolysis of RNA molecules into smaller components. They are divided into endonucleases (can cut RNA molecules in the middle) and exonucleases (degrades RNA from the extremities of the molecules).

Target

In gene expression profiling analysis, a target is the RNA transcript that is under investigation using its complementary counterpart, the probe.

Tissue Microarrays

Tissue microarrays (TMA) consist of paraffin blocks in which can be embedded with up to 1000 separate tissue cores, assembled in array fashion to allow simultaneous histological analysis.

Transcription

Transcription is the process by which DNA sequences are copied into complementary RNA molecules by the enzyme RNA polymerase. This reaction represents the transfer of genetic information from DNA into RNA, which is from “storing” to “function.” The DNA sequence that is transcribed into an RNA molecule is called a “transcript.”

Transcriptome

The transcriptome is the set of all RNA molecules, or “transcripts,” produced in one or a population of cells. The term can be applied to the total set of transcripts in a given organism, or to the specific subset of transcripts present in a particular cell type. Unlike the genome, which is roughly fixed for a given cell line (excluding mutations), the transcriptome can vary from cell to cell, and with external environmental conditions. Because it includes all RNA transcripts in the cell, the transcriptome reflects the genes that are being actively expressed at any given time. The study of the transcriptome examines the expression level of RNAs in a given cell population, often using high-throughput techniques based on DNA microarray technology, or RT-PCR.

Appendix C: Description of Genes

ONCO TYPE™: the 21-gene signature

| Acc | UGCluster | Name | Symbol | EGID | UGRepAcc | LLRepProtAcc | Chromosome | Cytoband |
|-----------|-----------|--|--------|-------|--------------|--------------|------------|---------------------|
| NM_001101 | Hs.520640 | Actin, beta | ACTB | 60 | AK125561 | NP_001092 | 7 | 7p15-p12 |
| NM_002046 | Hs.544577 | Glyceraldehyde-3-phosphate dehydrogenase | GAPDH | 2597 | BF983396 | NP_002037 | 12 | 12p13 |
| NM_001002 | Hs.546285 | Ribosomal protein, large, P0 | RPLP0 | 6175 | BQ051850 | NP_444505 | 12 | 12q24.2 |
| NM_000181 | Hs.255230 | Glucuronidase, beta | GUSB | 2990 | AK096764 | NP_000172 | 7 | 7q21.11 |
| NM_003234 | Hs.529618 | Transferrin receptor (p90, CD71) | TFRC | 7037 | BC001188 | NP_003225 | 3 | 3q29 |
| NM_002417 | Hs.80976 | Antigen identified by monoclonal antibody Ki-67 | MKI67 | 4288 | NM_002417 | NP_002408 | 10 | 10q25-qter |
| NM_003600 | Hs.250822 | Aurora kinase A | AURKA | 6790 | NM_198433 | NP_940839 | 20 | 20q13.2-q13.3 |
| NM_001168 | Hs.514527 | Effector cell peptidase receptor 1 | EPR1 | 8475 | NM_001012271 | | 17 | 17q25 |
| NM_031966 | Hs.23960 | Cyclin B1 | CCNB1 | 891 | NM_031966 | NP_114172 | 5 | 5q12 |
| NM_002466 | Hs.179718 | V-myb myeloblastosis viral oncogene homolog (avian)-like 2 | MYBL2 | 4605 | BX647151 | NP_002457 | 20 | 20q13.1 |
| NM_004448 | Hs.446352 | V-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian) | ERBB2 | 2064 | NM_001005862 | NP_004439 | 17 | 17q11.2-q12 17q21.1 |
| NM_005310 | Hs.86859 | Growth factor receptor-bound protein 7 | GRB7 | 2886 | NM_005310 | NP_005301 | 17 | 17q12 |
| NM_000043 | Hs.244139 | Fas (TNF receptor superfamily, member 6) | FAS | 355 | AB209361 | NP_690616 | 10 | 10q24.1 |
| NM_000926 | Hs.368072 | Progesterone receptor | PGR | 5241 | X51730 | NP_000917 | 11 | 11q22-q23 |
| NM_000633 | Hs.150749 | B-cell CLL/lymphoma 2 | BCL2 | 596 | NM_000633 | NP_000648 | 18 | 18q21.33 18q21.3 |
| NM_020974 | Hs.523468 | Signal peptide, CUB domain, EGF-like 2 | SCUBE2 | 57758 | NM_020974 | NP_066025 | 11 | 11p15.3 |
| NM_005940 | Hs.143751 | Matrix metalloproteinase 11 (stromelysin 3) | MMP11 | 4320 | NM_005940 | NP_005931 | 22 | 22q11.2 22q11.23 |
| NM_001333 | Hs.660866 | Cathepsin L2 | CTSL2 | 1515 | BC067289 | NP_001324 | 9 | 9q22.2 |
| NM_000561 | Hs.301961 | Glutathione S-transferase M1 | GSTM1 | 2944 | BQ880398 | NP_666533 | 1 | 1p13.3 |

| | | | | | | | | |
|-----------|-----------|----------------------------|------|-----|-----------|-----------|----|-------|
| NM_001251 | Hs.647419 | CD68 molecule | CD68 | 968 | NM_001251 | NP_001242 | 17 | 17p13 |
| NM_004323 | Hs.377484 | BCL2-associated athanogene | BAG1 | 573 | NM_004323 | NP_004314 | 9 | 9p12 |

MAMMAPRINT®: the 70-gene signature

| ORIGINALID | Acc | UGCluster | Name | Symbol | LLID | UGRepAcc | LLRepProtAcc | Chromosome | Cytoband |
|----------------|-----------|-----------|---|-------------|--------|--------------|--------------|------------|---------------|
| AA555029_RC | AA555029 | Hs.100691 | Hypothetical protein LOC286052 | LOC286052 | 286052 | AK095104 | | 8 | 8q24.13 |
| AF052162 | AF052162 | Hs.368853 | Acyltransferase like 2 | AYTL2 | 79888 | AK090444 | NP_079106 | 5 | 5p15.33 |
| NM_007203 | NM_007203 | Hs.591908 | PALM2-AKAP2 protein | PALM2-AKAP2 | 445815 | NM_053016 | NP_671492 | 9 | 9q31-q33 |
| AL080059 | AL080059 | Hs.173094 | TSPY-like 5 | TSPYL5 | 85453 | NM_033512 | NP_277047 | 8 | 8q22.1 |
| AL137718 | AL137718 | Hs.283127 | Diaphanous homolog 3 (Drosophila) | DIAPH3 | 81624 | NM_001042517 | NP_112194 | 13 | 13q21.2 |
| NM_003748 | NM_003748 | Hs.77448 | Aldehyde dehydrogenase 4 family, member A1 | ALDH4A1 | 8659 | NM_003748 | NP_733844 | 1 | 1p36 |
| NM_001282 | NM_001282 | Hs.514819 | Adaptor-related protein complex 2, beta 1 subunit | AP2B1 | 163 | NM_001030006 | NP_001273 | 17 | 17q11.2-q12 |
| U82987 | U82987 | Hs.467020 | BCL2 binding component 3 | BBC3 | 27113 | AF332558 | NP_055232 | 19 | 19q13.3-q13.4 |
| NM_004702 | NM_004702 | Hs.567387 | Cyclin E2 | CCNE2 | 9134 | NM_057735 | NP_477097 | 8 | 8q22.1 |
| NM_020974 | NM_020974 | Hs.523468 | Signal peptide, CUB domain, EGF-like 2 | SCUBE2 | 57758 | NM_020974 | NP_066025 | 11 | 11p15.3 |
| NM_001809 | NM_001809 | Hs.1594 | Centromere protein A | CENPA | 1058 | BM911202 | NP_001800 | 2 | 2p24-p21 |
| AF201951 | AF201951 | Hs.530735 | Membrane-spanning 4-domains, subfamily A, member 7 | MS4A7 | 58475 | NM_032597 | NP_996823 | 11 | 11q12 |
| X05610 | X05610 | Hs.508716 | Collagen, type IV, alpha 2 | COL4A2 | 1284 | NM_001846 | NP_001837 | 13 | 13q34 |
| Contig20217_RC | AA834945 | Hs.604604 | Transcribed locus, moderately similar to XP_001091104.1 similar to lin-9 homolog [Macaca mulatta] | AA834945 | | AA834945 | | 1 | |
| Contig24252_RC | AW024884 | Hs.528605 | PP12104 | LOC643008 | 643008 | XM_928053 | | 17 | 17q25.1 |
| Contig28552_RC | AA992378 | Hs.283127 | Diaphanous homolog 3 (Drosophila) | DIAPH3 | 81624 | NM_001042517 | NP_112194 | 13 | 13q21.2 |
| Contig32125_RC | AA404325 | Hs.523036 | CDNA FLJ38245 fis, clone FCBBF2007186 | AA404325 | | AK095564 | | 1 | |
| Contig32185_RC | AI377418 | Hs.657472 | G protein-coupled receptor 180 | GPR180 | 160897 | NM_180989 | NP_851320 | 13 | 13q32.1 |
| Contig35251_RC | AI283268 | Hs.634333 | CDNA: FLJ22719 fis, clone HSI14307 | AI283268 | | AK026372 | | 7 | |
| Contig38288_RC | AI554061 | Hs.657864 | Quiescin Q6 sulfhydryl oxidase 2 | QSOX2 | 169714 | AJ318051 | NP_859052 | 9 | 9q34.3 |

MAMMAPRINT®: the 70-gene signature (continued)

| ORIGINALID | Acc | UGCluster | Name | Symbol | LLID | UGRepAcc | LLRepProtAcc | Chromosome | Cytoband |
|----------------|-----------|-----------|--|----------|--------|--------------|--------------|------------|----------------|
| Contig40831_RC | AI224578 | Hs.595493 | Full-length cDNA clone CSODI029YM01 of Placenta Cot 25-normalized of Homo sapiens (human) | AI224578 | | BF675485 | | 8 | |
| Contig46218_RC | AI813331 | Hs.283127 | Diaphanous homolog 3 (Drosophila) | DIAPH3 | 81624 | NM_001042517 | NP_112194 | 13 | 13q21.2 |
| Contig46223_RC | AA528243 | Hs.22917 | Reticulon 4 receptor-like 1 | RTN4RL1 | 146760 | NM_178568 | NP_848663 | 17 | 17p13.3 |
| Contig48328_RC | AI694320 | Hs.655005 | Zinc finger protein 533 | ZNF533 | 151126 | BC092423 | NP_689733 | 2 | 2q31.2-q31.3 |
| Contig51464_RC | AI817737 | Hs.567582 | F-box protein 31 | FBXO31 | 79791 | AF318348 | NP_079011 | 16 | 16q24.2 |
| Contig55377_RC | AI918032 | Hs.632255 | RUN domain containing 1 | RUNDC1 | 146923 | BC039247 | NP_775102 | 17 | 17q21.31 |
| Contig55725_RC | AI992158 | Hs.470654 | Cell division cycle associated 7 | CDCA7 | 83879 | AL834186 | NP_665809 | 2 | 2q31 |
| Contig56457_RC | AI741117 | Hs.530272 | Chromosome 9 open reading frame 30 | C9orf30 | 91283 | AK092292 | NP_542386 | 9 | 9q31.1 |
| Contig63102_RC | AI583960 | Hs.55918 | Likely ortholog of mouse D11lgp2 | LGP2 | 79132 | AK021416 | NP_077024 | 17 | 17q21.2 |
| Contig63649_RC | AW014921 | Hs.446388 | CDNA FLJ41489 fis, clone BRTHA2004582 | AW014921 | | AK123483 | | 11 | |
| NM_020188 | NM_020188 | Hs.388255 | Chromosome 16 open reading frame 61 | C16orf61 | 56942 | BM463756 | NP_064573 | 16 | 16q23.2 |
| NM_000788 | NM_000788 | Hs.709 | Deoxycytidine kinase | DCK | 1633 | CD014015 | NP_000779 | 4 | 4q13.3-q21.1 |
| AL080079 | AL080079 | Hs.318894 | G protein-coupled receptor 126 | GPR126 | 57211 | NM_020455 | NP_940971 | 6 | 6q24.1 |
| Contig25991 | AI738508 | Hs.518299 | Epithelial cell transforming sequence 2 oncogene | ECT2 | 1894 | AY376439 | NP_060568 | 3 | 3q26.1-q26.2 |
| NM_007036 | NM_007036 | Hs.129944 | Endothelial cell-specific molecule 1 | ESM1 | 11082 | X89426 | NP_008967 | 5 | 5q11.2 |
| NM_000127 | NM_000127 | Hs.492618 | Exostoses (multiple) 1 | EXT1 | 2131 | NM_000127 | NP_000118 | 8 | 8q24.11-q24.13 |
| NM_003862 | NM_003862 | Hs.87191 | Fibroblast growth factor 18 | FGF18 | 8817 | AF075292 | NP_387498 | 5 | 5q34 |
| NM_018354 | NM_018354 | Hs.516834 | Chromosome 20 open reading frame 46 | C20orf46 | 55321 | AK126837 | NP_060824 | 20 | 20p13 |
| NM_002019 | NM_002019 | Hs.654360 | Fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor) | FLT1 | 2321 | NM_002019 | NP_002010 | 13 | 13q12 |
| NM_003875 | NM_003875 | Hs.591314 | Guanine monophosphate synthetase | GMPS | 8833 | NM_003875 | NP_003866 | 3 | 3q24 |

MAMMAPRINT®: the 70-gene signature (continued)

| ORIGINALID | Acc | UGCluster | Name | Symbol | LLID | UGRepAcc | LLRepProtAcc | Chromosome | Cytoband |
|------------|-----------|-----------|--|--------|-------|-----------|--------------|------------|---------------|
| NM_002073 | NM_002073 | Hs.584760 | Guanine nucleotide binding protein (G protein), alpha z polypeptide | GNAZ | 2781 | BC037333 | NP_002064 | 22 | 22q11.22 |
| NM_000849 | NM_000849 | Hs.2006 | Glutathione S-transferase M3 (brain) | GSTM3 | 2947 | NM_000849 | NP_000840 | 1 | 1p13.3 |
| NM_006101 | NM_006101 | Hs.414407 | NDC80 homolog, kinetochore complex component (S. cerevisiae) | NDC80 | 10403 | NM_006101 | NP_006092 | 18 | 18p11.32 |
| NM_018401 | NM_018401 | Hs.133062 | Serine/threonine kinase 32B | STK32B | 55351 | AY358353 | NP_060871 | 4 | 4p16.2-p16.1 |
| AF055033 | AF055033 | Hs.635441 | Insulin-like growth factor binding protein 5 | IGFBP5 | 3488 | NM_000599 | NP_000590 | 2 | 2q33-q36 |
| NM_000599 | NM_000599 | Hs.635441 | Insulin-like growth factor binding protein 5 | IGFBP5 | 3488 | NM_000599 | NP_000590 | 2 | 2q33-q36 |
| NM_014791 | NM_014791 | Hs.184339 | Maternal embryonic leucine zipper kinase | MELK | 9833 | NM_014791 | NP_055606 | 9 | 9p13.2 |
| AK000745 | AK000745 | Hs.377155 | Metadherin | MTDH | 92140 | BC045642 | NP_848927 | 8 | 8q22.1 |
| AB037863 | AB037863 | Hs.471955 | Early B-cell factor 4 | EBF4 | 57593 | XM_044921 | | 20 | 20p13 |
| NM_016448 | NM_016448 | Hs.656473 | Denticleless homolog (Drosophila) | DTL | 51514 | NM_016448 | NP_057532 | 1 | 1q32.1-q32.2 |
| NM_016359 | NM_016359 | Hs.615092 | Nucleolar and spindle associated protein 1 | NUSAP1 | 51203 | AK222819 | NP_060924 | 15 | 15q15.1 |
| NM_020386 | NM_020386 | Hs.36761 | HRAS-like suppressor | HRASLS | 57110 | BC048095 | NP_065119 | 3 | 3q29 |
| NM_005915 | NM_005915 | Hs.444118 | Minichromosome maintenance complex component 6 | MCM6 | 4175 | NM_005915 | NP_005906 | 2 | 2q21 |
| NM_004994 | NM_004994 | Hs.297413 | Matrix metalloproteinase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase) | MMP9 | 4318 | NM_004994 | NP_004985 | 20 | 20q11.2-q13.1 |
| NM_014889 | NM_014889 | Hs.528300 | Pitriysin metalloproteinase 1 | PITRM1 | 10531 | CR749279 | NP_055704 | 10 | 10p15.2 |
| NM_006681 | NM_006681 | Hs.418367 | Neuromedin U | NMU | 10874 | BF034907 | NP_006672 | 4 | 4q12 |
| NM_014321 | NM_014321 | Hs.49760 | Origin recognition complex, subunit 6 like (yeast) | ORC6L | 23594 | NM_014321 | NP_055136 | 16 | 16q12 |
| NM_000436 | NM_000436 | Hs.278277 | 3-oxoacid CoA transferase 1 | OXCT1 | 5019 | NM_000436 | NP_000427 | 5 | 5p13.1 |
| NM_006117 | NM_006117 | Hs.15250 | Peroxisomal D3,D2-enoyl-CoA isomerase | PECI | 10455 | AB209917 | NP_996667 | 6 | 6p24.3 |
| AF257175 | AF257175 | Hs.15250 | Peroxisomal D3,D2-enoyl-CoA isomerase | PECI | 10455 | AB209917 | NP_996667 | 6 | 6p24.3 |

MAMMAPRINT®: the 70-gene signature (continued)

| ORIGINALID | Acc | UGCluster | Name | Symbol | LLID | UGRepAcc | LLRepProtAcc | Chromosome | Cytoband |
|---------------|-----------|-----------|---|----------|-------|-----------|--------------|------------|--------------|
| NM_003607 | NM_003607 | Hs.35433 | CDC42 binding protein kinase alpha (DMPK-like) | CDC42BPA | 8476 | NM_003607 | NP_055641 | 1 | 1q42.11 |
| NM_003981 | NM_003981 | Hs.567385 | Protein regulator of cytokinesis 1 | PRC1 | 9055 | NM_003981 | NP_955446 | 15 | 15q26.1 |
| NM_016577 | NM_016577 | Hs.12152 | RAB6A, member RAS oncogene family | RAB6A | 5870 | NM_016577 | NP_942599 | 3 | 11q13.3 |
| NM_002916 | NM_002916 | Hs.518475 | Replication factor C (activator 1) 4, 37kDa | RFC4 | 5984 | NM_002916 | NP_853551 | 3 | 3q27 |
| AF073519 | AF073519 | Hs.658079 | Small EDRK-rich factor 1A (telomeric) | SERF1A | 8293 | AF073519 | NP_068802 | 5 | 5q12.2-q13.3 |
| NM_006931 | NM_006931 | Hs.419240 | Solute carrier family 2 (facilitated glucose transporter), member 3 | SLC2A3 | 6515 | AB209607 | NP_008862 | 12 | 12p13.3 |
| Contig2399_RC | W90004 | Hs.444450 | Egl nine homolog 1 (C. elegans) | EGLN1 | 54583 | AF229245 | NP_071334 | 1 | 1q42.1 |
| NM_003239 | NM_003239 | Hs.592317 | Transforming growth factor, beta 3 | TGFB3 | 7043 | AK122902 | NP_003230 | 14 | 14q24 |
| NM_015984 | NM_015984 | Hs.591458 | Ubiquitin carboxyl-terminal hydrolase L5 | UCHL5 | 51377 | AK225794 | NP_057068 | 1 | 1q32 |
| NM_003882 | NM_003882 | Hs.492974 | WNT1 inducible signaling pathway protein 1 | WISP1 | 8840 | AF100779 | NP_543028 | 8 | 8q24.1-q24.3 |

BCP – H/I assay: 2-gene signature and normalizing genes

| Symbol | UGCluster | Name | EGID | UGRepAcc | LLRepProtAcc | Chromosome | Cytoband |
|---------------|------------------|---|-------------|-----------------|---------------------|-------------------|-----------------|
| ACTB | Hs.520640 | Actin, beta | 60 | AK125561 | NP_001092 | 7 | 7p15-p12 |
| HMBS | Hs.82609 | Hydroxymethylbilane synthase | 3145 | BU168137 | NP_000181 | 11 | 11q23.3 |
| SDHA | Hs.440475 | Succinate dehydrogenase complex, subunit A, flavoprotein (Fp) | 6389 | AK131478 | NP_004159 | 5 | 5p15 |
| UBC | Hs.520348 | Ubiquitin C | 7316 | AB209436 | NP_066289 | 12 | 12q24.3 |
| HOXB13 | Hs.66731 | Homeobox B13 | 10481 | AY937237 | NP_006352 | 17 | 17q21.2 |
| IL17RB | Hs.654970 | Interleukin 17 receptor B | 55540 | NM_018725 | NP_758434 | 3 | 3p21.1 |

Symbol: official gene symbol

UGCluster: Unigene cluster identifier

Name: gene name according to Unigene

EGID: Entrez Gene identifier

UGRepAcc: representative GeneBank accession number according to Unigene

LLRepProtAcc: representative Protein accession number according to Entrez Gene

Chromosome: chromosomal location

Cytoband: cytogenetic band

Appendix D: Technologies

Reverse Transcription-Polymerase Chain Reaction (RT-PCR) and Real-Time RT-PCR

Reverse transcription polymerase chain reaction (RT-PCR) is a molecular biology technique for amplifying a specific piece of a ribonucleic acid (RNA) molecule. The RNA molecule is first reverse transcribed into complementary DNA (cDNA), followed by amplification of the resulting DNA by polymerase chain reaction (PCR), which is the common method used to amplify specific parts of a DNA molecule, via the temperature-mediated enzyme DNA polymerase. PCR uses specific short oligonucleotides, defined as primers, complementary to the target sequence to be amplified that serve to prime the polymerase reaction. The sequence of such oligonucleotides is responsible for the specificity of the reaction for the target nucleic acid fragment under analysis. PCR proceeds through subsequent amplification cycles determined by controlled temperature shifts of the reaction mixture. Real-time polymerase chain reaction is a laboratory technique that allows amplifying and quantifying simultaneously the specific part of the nucleic acid sequence under analysis. In this technique, the DNA quantity produced after each round of amplification is obtained by alternative methods. The most common quantification protocols are based on the use of fluorescent dyes that intercalate with double-strand DNA, or on modified DNA oligonucleotide probes that fluoresce when hybridized with the complementary DNA.

Real-time RT-PCR is the combination of the described techniques and enables gene expression evaluation at a particular time, or in a particular cell or tissue type. This technique is extremely sensitive and has been used to measure RNA from a single cell. The development of novel chemistries and instrumentation platforms has led to widespread use of this approach to measure gene expression changes. Moreover, this technique has become the preferred way to validate results obtained from microarray analyses and other techniques that evaluate gene expression changes on a global scale.

RT-PCR procedures

During PCR amplification, template, product and reagent relative ratios vary. At the beginning of the reaction, reagents are in excess, template and products are at low concentrations. In this phase they do not compete for primer binding, so that the amplification proceeds at an exponential rate. Following this initial phase the reaction enters a linear phase of amplification, in which annealing of the PCR products competes with primers for binding. Following this phase, in late reaction cycles, the amplification reaches a plateau and no more PCR products accumulate. Accurate and precise quantitative data are collected during the exponential phase of the amplification, in which amplification is extremely reproducible. In real-time PCR this process is automated and measurements are made at each cycle.

Several options are currently available to perform RT-PCR and real time RT-PCR: TaqMan[®] (Applied Biosystems, Foster City, CA, USA), Molecular Beacons, Scorpions[®] and the use of SYBR[®] Green (Molecular Probes). In all of these technologies PCR products are detected by generation of a fluorescent signal. TaqMan[®] probes, Molecular Beacons and Scorpions[®] rely on Förster Resonance Energy Transfer (FRET): a dye molecule and a quencher moiety are bound to the same or different oligonucleotide substrates and fluorescence is emitted when they are separated. SYBR Green is a fluorogenic dye that emits a strong fluorescent signal upon binding to double-stranded DNA.

TaqMan probes

TaqMan technology depends on the 5'-nuclease activity of the DNA polymerase used for PCR. This activity is used to separate the quencher and the dye, releasing FRET and thus producing fluorescence. During the reaction, this enzyme hydrolyzes the oligonucleotide probes that are hybridized to the target

sequence, decoupling occurs, and fluorescence arises, increasing at each cycle, proportional to the amount of probe cleavage.

Molecular beacons

Molecular Beacons also is based on FRET, although the design of the probes is different. In this chemistry, a dye is attached to the 5' end and a quencher is bound to the 3' end of an oligonucleotide substrate. The 5'-nuclease activity of the DNA polymerase is not required since Molecular Beacons probes in solution form a loop structure that prevents fluorescing, while after hybridization to the target sequence, the dye and quencher are separated, FRET is release, and light is emitted upon irradiation.

Scorpions

Scorpion technology assembles the amplification primer and the reporter sequence into the same oligonucleotide. In solution, the dye is attached to the 5' end of the probe and is quenched by a moiety coupled to a complementary sequence, linked to the primer at the 3' end, through a non-amplifiable monomer. During PCR, after extension of the Scorpion primer, the two specific probe sequences are able to bind each other, thus opening up the hairpin loop, releasing quenching and causing signal emission.

SYBR Green

SYBR Green binds double-stranded DNA, and upon excitation fluoresces. The more PCR products accumulate, the more light emission increases. SYBR® Green is sensitive, inexpensive, and easy to use. However it binds to any double-stranded DNA molecule in the reaction, including primer-dimers and other non-specific reaction products, and this may result in an overestimation of the target molecule concentration. Since this dye binds to double-stranded DNA, there is no need to design specific probes for any particular target under analysis.

Real-time reporters for Multiplex PCR

Several implementations of this technique (TaqMan, Molecular Beacons and Scorpions) allow multiple DNA species to be measured in the same sample (multiplex PCR). Fluorescent dyes with different emission spectra, indeed, may be coupled to the different probes assaying different targets. This approach allows the use of internal controls, which can be co-amplified along with the target sequence under analysis in the same reaction tube. Multiplex is not possible with SYBR Green.

Quantitation of results

Two methods are commonly used to quantify the results obtained by real-time RT-PCR:

1. The standard curve method;
2. Comparative threshold method;

The standard curve method

In this method, a standard curve is obtained from a nucleic acid template of known concentration, serially diluted. This curve is subsequently used as a reference to extrapolate quantitative information about mRNA targets of unknown concentrations. Such standards can be RNA molecules transcribed in vitro from cDNA plasmids, or other nucleic acid templates prepared at the purpose. cDNA plasmids are the preferred standards used to obtain the standard curve, however, their use will not allow inferences about the efficiency of the reverse transcription reaction, or about possible differences in the RNA template inputs. For this reason normalization to one or more housekeeping genes is often used.

Comparative threshold method

This approach involves the comparison of the cycle threshold (CT) values of the samples of interest to the CT values of a control RNA sample, after internal normalization of each CT to an appropriate endogenous housekeeping gene. For this method to be valid, the amplification efficiencies of the target

and the endogenous reference must be similar. If a housekeeping gene cannot be found, whose amplification efficiency is similar to the target, then the standard curve method is better.

Instrumentation for Real-time RT-PCR

Real-time RT-PCR requires platforms consistency of a thermal cycler, a computer, optics for fluorescence excitation and emission collection, and data acquisition and analysis software. Such instrumentation, available from several manufacturers, varies in term of sample capacity (single tubes, 96-well, 384-well formats), excitation method, and overall sensitivity.

DNA Microarrays

The introduction of automated large scale sequencing, supported by adequate computational tools and bioinformatics development, has greatly increased our general knowledge on genomic sequences organization and function. This knowledge is the basis for gene expression investigation on a global scale by parallel analysis of thousands of genes in a single assay. In microarray analysis, the Northern blotting scheme is reversed: the labeled moiety is obtained from the RNA sample and a certain number of immobilized known sequences are used as probes (Baldwin, Crane et al. 1999). The advances made in attaching nucleic acid sequences to glass supports and robotics allowed investigators to miniaturize the scale of the reactions. Modified microscope slides could be used to deposit thousands of nucleic acid sequences. The same result was also obtained by borrowing photolithography techniques from semiconductor manufacturing to synthesize oligonucleotides directly onto a solid support (Watson, Mazumder et al. 1998). Altogether, these advances led to in 1995, to the first papers in which the term “microarray” was used in its current meaning (Schena, Heller et al. 1998).

Principles of microarray analysis

All the different technical solutions that have been so far developed to perform microarray analysis are miniaturized hybridization assays that allow investigators to simultaneously query thousands of nucleic acid fragments. All microarray systems share the following key components:

- The array, which contains the immobilized nucleic acid sequences, known as “probes”;
- One or more labeled samples or “targets”, that are hybridized against the microarray;
- A detection system that quantify the hybridization signals

Microarrays and DNA-chips

Spotted microarrays consist of a collection of preformed nucleic acid sequences immobilized onto the solid support so that each unique sequence forms a tiny feature called “spot” or feature. These nucleic acids are obtained in numerous ways, and there are different methods for depositing them onto microarray slides (by simple contact, by ink-jet technology, or by micro-syringe pumping for instance). In general, nucleic acid prepared for deposition on microarrays consist of cDNA clones amplified by polymerase chain reaction (cDNA microarrays), or of synthesized oligonucleotides of various length (oligonucleotide microarrays, i.e., microarrays from Agilent Technologies). The size of the spots differs from one system to another, but it is usually less than two hundred micrometers in diameter. A modified glass slide or glass wafer acts as the solid support onto which up to tens of thousands of spots can be arrayed in a total area of a few square centimetres. On the contrary, DNA-chips are produced by a proprietary technology (GeneChip®, Affymetrix) quite different from the spotted one, as it is based on

direct photolithography synthesis of short oligonucleotides (20-25 base pairs) on the solid support.

Target labeling

Whatever the kind of microarray used, DNA probes present on the arrays are interrogated by nucleic acid hybridization with a labeled target. The sample may be mRNA for a gene expression study or genomic DNA for other purposes (promoter usage analysis: CHIP-on-Chip, genomic rearrangements: FISH-on Chip). The sample is converted to a labeled population of nucleic acids, known as the target. These moieties consist of several thousands of different labeled nucleic acid fragments and the final complexity is much greater than the one usually encountered in other routine molecular biology experiments. Therefore, these hybridizations should be carried out under conditions that do not promote annealing of non-complementary fragments. Fluorescent dyes, and especially the cyanine dyes Cy3 and Cy5, have been widely adopted as the predominant labels in microarray analysis. Fluorescence has the advantage of permitting the detection of two or more different signals in one experiment. This has thus allowed investigators to perform comparative analysis of two or more samples on one microarray. The described scheme is usually adopted for cDNA microarray analysis, while single channel experiments are the best-suited choice for GeneChip® technology, thanks to the high manufacturing reproducibility of the chips. The use of fluorescence has also increased the accuracy and throughput of microarray analysis over filter-based macroarrays, in which only one radioactively labeled sample can be conveniently analyzed at a time.

Microarray hybridization

In a microarray hybridization, the labeled fragments in the target are expected to form duplexes with their immobilized complementary probes. This requires that the nucleic acids are single-stranded and accessible to each other. The number of duplexes formed reflects the relative number of each specific fragment in the target, as long as the amount of immobilized nucleic acid probe is in excess and not restraining the kinetics of hybridization. Two or more samples labeled with different fluorescent dyes can be hybridized simultaneously, resulting in simultaneous hybridization taking place at each spot. By measuring the different fluorescent signals associated with each feature, the relative abundance of specific sequences in each of the samples can be determined.

Scanning and data analysis

Microarray scanners typically contain two different lasers that emit light at wavelengths that are suitable for exciting the fluorescent dyes used as labels. A detector system attached to a confocal microscope records the emitted light from each feature of the array, permitting high-resolution detection of the hybridization signals. Alternative solutions use CCD-camera devices to detect the fluorescence. Despite their small size, microarrays allow the generation of a large amount of data even from a single hybridization. For these reasons the use of computerized data processing is necessary in order to handle the amount of generated data and to gain maximum information from the experiment. This is usually achieved by specialized software that extracts primary data from scanned microarray slide images, normalizes this data to remove the influence of experimental variation, and finally manipulates the data so that biologically meaningful conclusions can be made.

Applications of microarray analysis

The versatility of microarray analysis is confirmed by its rapid emergence as a general molecular biology analytical technique. An increasing number of researchers are now exploiting this technology in diverse biomedical disciplines. In fact, microarrays have not replaced established techniques, but rather represent a high-power approach to perform analyzes that were previously time consuming. By using information derived from the several complete or near complete genome sequences, including the human genome, it is now possible to perform genome-wide experiments using microarray technology. This has already been demonstrated for *Saccaromyces Cerevisiae* where all the expressed genes are known (Chu, DeRisi et al. 1998; Spellman, Sherlock et al. 1998). Due to the availability of millions of data points at once, microarrays enabled global analysis of fundamental biological processes: gene expression analysis, genome analysis, and drug discovery have been three of the main areas in which microarray analysis has been applied so far.

Gene expression analysis

Gene expression analysis examines the composition of cellular messenger RNA populations. The identity of transcripts that make up these populations and their expression levels are informative of the cell state and of the activity of the genes and, as the precursors of translated proteins, changes in mRNA levels are related to changes in the proteome. In the simplest scheme a typical microarray gene expression experiment compares the relative expression levels of specific transcripts in two samples. Usually one of the samples is a control while the other is obtained from cells whose response or status is being explored. Each one of the two samples is labeled with a different fluorescent dye, and equal amounts of the labeled samples are combined and hybridized with the microarray. After hybridization, two grey scale images (usually in a 16-bit TIFF format) corresponding to the fluorescent signals of the two dyes are independently obtained by scanning the microarray and fluorescence intensity from each feature is subsequently quantified by a specific software. After normalization, the intensity of the two hybridization signals can be compared: equal signal from both samples suggests equal expression of the considered genes in both samples, while signals' disparity is suggestive of differential expression.

One of the most important remarks that has to be taken into account is that microarray analysis does not give any information about absolute gene expression levels in the samples. This is because the intensity of the fluorescent signals is not only proportional to the number of hybridized fragments, but also to the length of these fragments and the number of fluorescent labels each fragment carries (specific activity of the target or labeling density). These parameters are determined by the unique nucleotide sequence of each transcript, so that they will vary from gene to gene. If the two samples have been labeled under similar conditions, the length and labeling density of specific transcripts will be similar, allowing the comparison of the relative abundance of the transcripts in the analyzed targets. For these reasons a strong hybridization signal from microarray analysis does not necessarily correspond to a highly expressed gene, as it could be derived, for instance, from a gene that is expressed at a relatively low level but yields highly labeled target fragments.

Gene expression analysis with microarrays has been applied to numerous mammalian tissues, plants, yeast, and bacteria (Braxton and Bedilion 1998; Mirnics 2001; Mirnics, Middleton et al. 2001; Schulze and Downward 2001; van Berkum and Holstege 2001). These studies have examined the effects of treating cells with chemicals, the consequences of over-expression of

regulatory factors in transfected cells, and compared mutant strains with parental strains to delineate functional pathways. In cancer research, microarrays have been used to find gene expression changes in transformed cells and metastases, to identify diagnostic markers, and to classify tumors based on their gene expression profiles (DeRisi, Penland et al. 1996; Alizadeh, Eisen et al. 1999; Alizadeh, Ross et al. 2001; Rew 2001, van't Veer et al. 2002).

Genomic analysis

In addition to gene expression analysis, microarrays are now also established tools for genomic analysis (Shoemaker, Schadt et al. 2001). Microarrays, in fact, can be used to reveal transcription factor interactions with specific sequences and motifs regulating gene expression. For example, by combining immunoprecipitation of transcription factor-DNA complexes to microarray identification of DNA fragments on a genomic microarray, it was possible to identify functional regulatory elements in the yeast genome (Lieb, Liu et al. 2001). Furthermore, microarrays were used to predict splice variants of transcripts and investigate genomic fragments derived from genetic analysis methods, such as genomic mismatch scanning and representational difference analysis (Hu, Madore et al. 2001; Meltzer 2001) and specific oligonucleotide microarrays have been applied to the analysis of known single nucleotide polymorphisms (SNPs) and mutations (Sapolsky, Hsie et al. 1999; Larsen, Christiansen et al. 2001). Moreover, microarray hybridization can also be used to sequence DNA samples, thus providing a suitable mean for identifying new genetic variants (Drobyshev, Mologina et al. 1997).

Drug discovery

A typical drug discovery process needs several years of research and only a few candidate compounds result at the end in approved drugs. For these reasons methods that increase the efficiency of the process and improve the probability of developing effective drugs are needed. In this perspective microarray analysis proved useful in different stages of drug discovery (Lockhart and Winzeler 2000; Meltzer 2001; van Berkum and Holstege 2001). For instance, the identification of potential therapeutic compounds can be achieved by elucidating metabolic pathways by looking for co-expressed. Once the drug candidates have been selected, microarrays can be subsequently used to define their toxic properties by examining expression profiles induced by drug treatments (Jain 2000). Moreover, the gene expression changes elicited by different drug treatments were also recently used to recognize their mechanisms of action (Jain 2000).

Specific Gene Expression Based Assays

***Oncotype DX*TM process**

Oncotype DX is a multi-gene assay, designed to provide a quantitative assessment of the likelihood of breast cancer distant recurrence. *Oncotype DX* is offered by Genomic Health, where the assay was developed. The assay accounts for the following procedures: RNA is extracted and purified from the tumor specimen, then the level of expression of 21 genes (16 cancer related and 5 control genes) is obtained by RT-PCR, finally the Recurrence ScoreTM is calculated from the gene expression results.

In the current implementation of the assay, a pathologist at Genomic Health reviews the tumor content of the specimens to be processed, then RNA is extracted from formalin-fixed, paraffin-embedded (FFPE) specimens and contaminant DNA is removed by DNase I treatment.

Total RNA yield is measured and the absence of DNA contamination is verified. Real time RT-PCR is performed by TaqMan® technology in 384-well plates. The expression of the 16 cancer genes is measured in triplicate then normalized to the expression levels of the 5 reference genes. Finally, normalized gene expression levels of the 16 cancer related genes are used to compute the Recurrence Score (RS), on a scale from 1 to 100. Clinical studies showed the correlation of the RS with the likelihood of distant recurrence at 10 years, which increases continuously with increase of the RS, however three distinct groups of risk were defined: low-risk (RS < 18), intermediate-risk (RS 18-30), and high-risk (RS ≥ 31) (Paik, Shak et al. 2004).

The *Oncotype DX* test is offered to patients who meet the following criteria:

- Newly diagnosed
- Will be treated with tamoxifen;
- Stage I invasive breast cancer with ER positive;
- Stage II invasive breast cancer with ER positive and lymph node negative.

MammaPrint®

MammaPrint is a multi-gene microarray-based, diagnostic assay, designed to provide a quantitative prediction of risk of metastasis in breast cancer patients. The assay measures in triplicate the expression levels of 70 distinct genes, which were originally identified in a research performed at the Netherlands Cancer Institute (Amsterdam, The Netherlands) (van 't Veer, Dai et al. 2002; van de Vijver, He et al. 2002). According to this test, patients are divided into two risk groups, with different prognosis, by measuring the cosine correlation between the 70-gene expression profile of each individual patient to the original signature developed, according to a pre-specified threshold.

The CE-marked, FDA cleared assay is offered by the certified (QSR/GMP, ISO 17025, CLIA (#99D1030869) and CAP) Agendia laboratory (Amsterdam, The Netherlands), with the following features:

- The assay is performed from fresh (non-FFPE) specimens;
- A validated sampling and transportation method of fresh tissue on ambient temperature;
- A histologic review of the shipped specimens;
- RNA extraction and quality evaluation prior to microarray analysis;
- Triplicate gene expression measurements and duplicate sample measurements, in a dye-swap design;
- Use of a constant, standardized reference RNA in each hybridization;

The MammaPrint test is offered to patients who meet the following criteria:

- Below age 61;
- Stage I invasive breast cancer with ER positive or ER negative;
- Stage II invasive breast cancer with ER positive or ER negative and lymph node negative;
- Tumor size less than 5 cm.

Breast Cancer Profiling (BCP or H/I ratio)

The Breast Cancer Profiling (BCP) assay is based on the two-gene expression index (HOXB13/IL17BR) developed by Ma and colleagues (Ma, Wang et al. 2004; Ma, Hilsenbeck et al. 2006). Gene expression levels for the two genes are measured by real time RT-PCR,

normalized to a specific set of reference genes, prior the index computation. The two-gene index is a continuous marker of recurrence risk in untreated ER-positive, node negative patients.

This assay is licensed by AviaraDX to Quest Diagnostic, and it is offered as a laboratory service, with the following features:

- The assay is performed from FFPE specimens;
- Laser capture microdissection is performed if the specimen content is <30% cancer cells;
- RNA preparation and quality evaluation;
- Real-time PCR analysis of HOXB13 and IL-17BR gene expression;
- Formulation of the normalized two-gene expression index;
- Result formulation with 5-year recurrence risk;

The BCP assay is offered to patients who meet the following criteria:

- Treatment-naïve individuals with ER-positive/lymph node-negative breast cancer

Bibliography

- Alizadeh, A., M. Eisen, et al. (1999). "The lymphochip: a specialized cDNA microarray for the genomic-scale analysis of gene expression in normal and malignant lymphocytes." *Cold Spring Harb Symp Quant Biol* 64: 71-8.
- Alizadeh, A. A., D. T. Ross, et al. (2001). "Towards a novel classification of human malignancies based on gene expression patterns." *J Pathol* 195(1): 41-52.
- Baldwin, D., V. Crane, et al. (1999). "A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants." *Curr Opin Plant Biol* 2(2): 96-103.
- Braxton, S. and T. Bedilion (1998). "The integration of microarray information in the drug development process." *Curr Opin Biotechnol* 9(6): 643-9.
- Chu, S., J. DeRisi, et al. (1998). "The transcriptional program of sporulation in budding yeast." *Science* 282(5389): 699-705.
- DeRisi, J., L. Penland, et al. (1996). "Use of a cDNA microarray to analyse gene expression patterns in human cancer." *Nat Genet* 14(4): 457-60.
- Drobyshev, A., N. Mologina, et al. (1997). "Sequence analysis by hybridization with oligonucleotide microchip: identification of beta-thalassemia mutations." *Gene* 188(1): 45-52.
- Hu, G. K., S. J. Madore, et al. (2001). "Predicting splice variant from DNA chip expression data." *Genome Res* 11(7): 1237-45.
- Jain, K. K. (2000). "Applications of biochip and microarray systems in pharmacogenomics." *Pharmacogenomics* 1(3): 289-307.
- Larsen, L. A., M. Christiansen, et al. (2001). "Recent developments in high-throughput mutation screening." *Pharmacogenomics* 2(4): 387-99.
- Lieb, J. D., X. Liu, et al. (2001). "Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association." *Nat Genet* 28(4): 327-34.
- Lockhart, D. J. and E. A. Winzler (2000). "Genomics, gene expression and DNA arrays." *Nature* 405(6788): 827-36.
- Ma, X. J., S. G. Hilsenbeck, et al. (2006). "The HOXB13:IL17BR expression index is a prognostic factor in early-stage breast cancer." *J Clin Oncol* 24(28): 4611-9.
- Ma, X. J., Z. Wang, et al. (2004). "A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen." *Cancer Cell* 5(6): 607-16.
- Meltzer, P. S. (2001). "Spotting the target: microarrays for disease gene discovery." *Curr Opin Genet Dev* 11(3): 258-63.
- Mirnics, K. (2001). "Microarrays in brain research: the good, the bad and the ugly." *Nat Rev Neurosci* 2(6): 444-7.
- Mirnics, K., F. A. Middleton, et al. (2001). "Analysis of complex brain disorders with gene expression microarrays: schizophrenia as a disease of the synapse." *Trends Neurosci* 24(8): 479-86.
- Paik, S., S. Shak, et al. (2004). "A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer." *N Engl J Med* 351(27): 2817-26.
- Rew, D. A. (2001). "DNA microarray technology in cancer research." *Eur J Surg Oncol* 27(5): 504-8.
- Sapolsky, R. J., L. Hsie, et al. (1999). "High-throughput polymorphism screening and genotyping with high-density oligonucleotide arrays." *Genet Anal* 14(5-6): 187-92.
- Schena, M., R. A. Heller, et al. (1998). "Microarrays: biotechnology's discovery platform for functional genomics." *Trends Biotechnol* 16(7): 301-6.
- Schulze, A. and J. Downward (2001). "Navigating gene expression using microarrays--a technology review." *Nat Cell Biol* 3(8): E190-5.
- Shoemaker, D. D., E. E. Schadt, et al. (2001). "Experimental annotation of the human genome using microarray technology." *Nature* 409(6822): 922-7.
- Spellman, P. T., G. Sherlock, et al. (1998). "Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization." *Mol Biol Cell* 9(12): 3273-97.
- van 't Veer, L. J., H. Dai, et al. (2002). "Gene expression profiling predicts clinical outcome of breast cancer." *Nature* 415(6871): 530-6.
- van Berkum, N. L. and F. C. Holstege (2001). "DNA microarrays: raising the profile." *Curr Opin Biotechnol* 12(1): 48-52.
- van de Vijver, M. J., Y. D. He, et al. (2002). "A gene-expression signature as a predictor of survival in breast cancer." *N Engl J Med* 347(25): 1999-2009.
- Watson, A., A. Mazumder, et al. (1998). "Technology for microarray analysis of gene expression." *Curr Opin Biotechnol* 9(6): 609-14.

Appendix E: Technical Experts and Peer Reviewers

Technical Experts and Peer Reviewers

Katrina Armstrong, M.D., M.S.C.E.*
Director of Research
Leonard Davis Institute of Health
Economics
University of Pennsylvania School of
Medicine
Philadelphia, PA

Richard A. Bender, M.D., F.A.C.P.
Medical Director-Oncology
Quest Diagnostics Nichols Institute-SJC
San Juan Capistrano, CA

Lyndsay N. Harris, M.D.
Assoc Prof Med
Dir Breast Cancer Disease Unit and Int Med
Medical Oncology
333 Cedar St
New Haven, CT

Daniel F. Hayes, M.D.
University of Michigan
Comprehensive Cancer Center
Ann Arbor, MI

Charles Perou, Ph.D.
University of North Carolina, Chapel Hill
Department of Genetics
Chapel Hill, NC

Kathryn Phillips, Ph.D.*
Prof. of Health Economics and Health
Services Research
School of Pharmacy, Institute for Health
Policy Studies, and UCSF Comprehensive
Cancer Center
University of California, San Francisco
San Francisco, CA

Margaret Piper, M.P.H., Ph.D.*
Associate Director
Blue Cross/Blue Shield Association
Technology Evaluation Center
Steve Teutsch, M.D., M.P.H.*
Executive Director of Outcomes Research
Merck & Co., Inc.
West Point, PA

Steven Shak, M.D.
Chief Medical Officer
Genomic Health
Redwood City, CA

Richard Simon, D.Sc.
NCI
Chief, Biometric Research Branch
Rockville, MD

Laura van't Veer, Ph.D.
Netherlands Cancer Institute
Amsterdam
The Netherlands

* Evaluation of Genomic Applications in Practice
and Prevention (EGAPP) working group member.

Internal Technical Experts

Giovanni Parmigianni, M.S., Ph.D.
Johns Hopkins University
School of Medicine, Oncology
Bioinformatics
550 Building, 11-03
Baltimore, MD

Antonio C. Wolf, M.D.
Johns Hopkins University
School of Medicine, Oncology Center
CRB 189
1650 Orleans Street
Baltimore, MD

Appendix F: Detailed Electronic Database Search Strategies

MEDLINE Strategy

| | |
|---|------|
| ((("breast neoplasms"[mh] OR "breast cancer"[tiab] OR (breast[tiab] AND neoplasm[tiab])) AND ((Gene[tiab] AND expression[tiab]) OR "gene expression profiling"[mh] OR "gene expression"[mh]) AND 1990:2007[dp] AND Eng[lang]) NOT((animals[mh]NOT humans[mh]) OR review[pt])) NOT Tumor Cells, Cultured[mh] | 3356 |
|---|------|

Cochrane Library (Reviews and CENTRAL) Strategy

| | |
|--|----|
| ("breast neoplasms" or "breast cancer":ti,ab,kw or (breast AND cancer):ti,ab,kw or (breast AND neoplasm):ti,ab,kw) AND ("gene expression profiling" or "gene expression" or "gene expression" AND profiling:ti,ab,kw or "gene expression" AND (test OR tesing):ab) | 55 |
|--|----|

EMBASE Strategy

| | |
|---|------|
| ((("breast tumor"/exp) OR (breast:ti,ab AND cancer:ti,ab)) AND (((('gene expression'/exp) OR ('gene expression profiling'/exp)) OR ('gene expression':ab,ti AND profiling:ab,ti))) NOT (('cell culture'/exp) OR ('validation study'/exp) OR (apoptosis:ab,ti) OR ('cell death':ab,ti) OR (transcriptional:ti,ab AND mechanism:ti,ab) OR (transcriptional:ti,ab AND machinery:ti,ab)) AND [english]/lim AND [humans]/lim AND [1990-2007]/py) NOT (review:it) | 7531 |
|---|------|

CINAHL

| | |
|---|----|
| ((MH "breast neoplasms" or TX "breast cancer") OR (TX (Breast AND cancer) or TX (breast AND neoplasm))) AND (TX "gene expression profiling" or TX ("gene expression" AND profiling) or TX (gene AND profiling)) | 73 |
|---|----|

MEDLINE (targeted authors search)

| | |
|--|------|
| ((van't veer LJ[au] OR Dai H[au] OR van de vijver MJ[au] OR He YD[au] OR Hart AM[au] OR Hart AA[au] OR Mao M[au] OR Peterse HL[au] OR van der kooy K[au] OR Marton MJ[au] OR Witteveen AT[au] OR Schreiber GJ[au] OR Kerkhoven RM[au] OR Roberts C[au] OR Linsley PS[au] OR Bernards R[au] OR Friend SH[au] OR Voskuil DW[au] OR Parrish M[au] OR Atsma D[au] OR Witteveen A[au] OR Glas A[au] OR Delahaye L[au] OR van der velde T[au] OR Bartelink H[au] OR Rodenhuis S[au] OR Rutgers ET[au]) OR (paik S[au] OR shak S[au] OR Tang G[au] OR Kim C[au] OR Baker J[au] OR Cronin M[au] OR baehner FL[au] OR walker MG[au] OR Watson D[au] OR Park T[au] OR Hiller W[au] OR Fisher ER[au] OR Wickerham DL[au] OR Bryant J[au] OR Wolmark N[au]) OR (Ma XJ[au] OR Wang Z[au] OR Ryan PD[au] OR Isakoff SJ[au] OR Barmettler A[au] OR Fuller A[au] OR Muir B[au] OR Mohapatra G[au] OR Salunga R[au] OR Tuggle JT[au] OR Tran Y[au] OR tran D[au] OR Tassin A[au] OR Amon P[au] OR Wang W[au] OR Enright E[au] OR Stecker K[au] OR Estepa-Sabal E[au] OR Smith B[au] OR Younger J[au] OR Balis U[au] OR Michaelson J[au] OR bhan A[au] OR Habin K[au] OR Baer TM[au] OR Brugge J[au] OR Haber AH[au] OR Erlander MG[au] OR Sgroi DC[au])) AND gene[tw] AND ((Gene[tiab] AND expression[tiab]) OR "gene expression profiling"[mh] OR "gene expression"[mh]) AND 1990 : 2007[dp] AND Eng[lang] NOT (animals[mh] NOT humans[mh]) OR review[pt]) | 1947 |
|--|------|

Appendix G: Title Review Form

1. **Record ID: 1781**

van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., and Friend, S. H. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415(6871):530-6.

Does this article POTENTIALLY apply to the key questions?

- POTENTIALLY eligible
- INELIGIBLE

Appendix G: Abstract Review Form

1. **Record ID: 1781**

van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., and Friend, S. H. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415(6871):530-6.

ABSTRACT: Breast cancer patients with the same stage of disease can have markedly different treatment responses and overall outcome. The strongest predictors for metastases (for example, lymph node status and histological grade) fail to classify accurately breast tumours according to their clinical behaviour. Chemotherapy or hormonal therapy reduces the risk of distant metastases by approximately one-third; however, 70-80% of patients receiving this treatment would have survived without it. None of the signatures of breast cancer gene expression reported to date allow for patient-tailored therapy strategies. Here we used DNA microarray analysis on primary breast tumours of 117 young patients, and applied supervised classification to identify a gene expression signature strongly predictive of a short interval to distant metastases ('poor prognosis' signature) in patients without tumour cells in local lymph nodes at diagnosis (lymph node negative). In addition, we established a signature that identifies tumours of BRCA1 carriers. The poor prognosis signature consists of genes regulating cell cycle, invasion, metastasis and angiogenesis. This gene expression profile will outperform all currently used clinical parameters in predicting disease outcome. Our findings provide a strategy to select patients who would benefit from adjuvant therapy.

Should this article be REVIEWED? (choose one)

[1] YES: indicate the questions that this article might apply to (below)

This article potentially applies to the following key questions (Choose all that apply)

1. What is the **direct** evidence that the Mammaprint or OncotypeDX gene expression profiling tests in women diagnosed with breast cancer (or any specific subset of this population) lead to improvement in outcomes?
2. What are the sources of and contributions to **analytic variability** in these two gene expression-based prognostic estimators for women diagnosed with breast cancer?
3. What is the **clinical validity** of these tests in women diagnosed with breast cancer?
 - a. How well does this testing predict recurrence rates for breast cancer compared to standard prognostic approaches? Specifically, how much do these tests add to currently known factors or combination indices that predict the probability of breast cancer recurrence, (e.g., tumor type or stage, age, estrogen receptor (ER) and human epidermal growth factor receptor 2 (HER2) status)?
 - b. Are there any other factors, which may not be components of standard predictors of recurrence (e.g., race/ethnicity or adjuvant therapy), that affect the clinical validity of these tests, and thereby generalizability of results to different populations?
4. What is the **clinical utility** of these tests?
 - a. To what degree do the results of these tests predict the response to chemotherapy, and what factors affect the generalizability of that prediction?
 - b. What are the effects of using these two tests and the subsequent management options on the following outcomes: testing or treatment related psychological harms, testing or treatment related physical harms, disease recurrence, mortality, utilization of adjuvant therapy, and medical costs.
 - c. What is known about the utilization of Mammaprint and OncotypeDX gene expression profiling in women diagnosed with breast cancer in the United States?
 - d. What projections have been made in published analyses about the cost-effectiveness of using Mammaprint and OncotypeDX gene expression profiling in women diagnosed with breast cancer?

[2] Unclear/No abstract (promote to article review)

[3] NOT eligible (exclude): indicate reason for exclusion (below)

Reason for EXCLUSION? (choose any that apply)

- [1] Study applies only to breast cancer biology
- [2] Study only applies to single or multiple gene predictors and does not involve OncotypeDX or Mammaprint profiles
- [3] Does not involve OncotypeDX or Mammaprint gene expression profiling tests
- [4] Does not involve original data or original data analysis
- [5] Does not involve women
- [6] Does not involve breast cancer patients
- [7] Not English language
- [8] Does not apply to the key questions
- [9] OTHER _____
- [10] Unclear

[4] No, may be useful for BACKGROUND material (pull for hand searching If publish in 2002 or later)

Appendix G: Article Review Form

ARTICLE inclusion/exclusion

Record ID: 750

Reid, J. F., Lusa, L., De Cecco, L., Coradini, D., Veneroni, S., Daidone, M. G., Gariboldi, M., and Pierotti, M. A. Limits of predictive models using microarray data for breast cancer clinical treatment outcome. *Journal of the National Cancer Institute* 2005;97(12):927-30.

ABSTRACT:

Should this article be REVIEWED? (choose one)

[1] **YES: indicate the questions that this article might apply to (below)**

This article potentially applies to the following key questions (Choose all that apply)

1. What is the **direct** evidence that the Mammprint or OncotypeDX gene expression profiling tests in women diagnosed with breast cancer (or any specific subset of this population) lead to improvement in outcomes?
2. What are the sources of and contributions to **analytic variability** in these two gene expression-based prognostic estimators for women diagnosed with breast cancer?
3. What is the **clinical validity** of these tests in women diagnosed with breast cancer?
 - a. How well does this testing predict recurrence rates for breast cancer compared to standard prognostic approaches? Specifically, how much do these tests add to currently known factors or combination indices that predict the probability of breast cancer recurrence, (e.g., tumor type or stage, age, estrogen receptor (ER) and human epidermal growth factor receptor 2 (HER2) status)?
 - b. Are there any other factors, which may not be components of standard predictors of recurrence (e.g., race/ethnicity or adjuvant therapy), that affect the clinical validity of these tests, and thereby generalizability of results to different populations?
4. What is the **clinical utility** of these tests?
 - a. To what degree do the results of these tests predict the response to chemotherapy, and what factors affect the generalizability of that prediction?
 - b. What are the effects of using these two tests and the subsequent management options on the following outcomes: testing or treatment related psychological harms, testing or treatment related physical harms, disease recurrence, mortality, utilization of adjuvant therapy, and medical costs.
 - c. What is known about the utilization of Mammprint and OncotypeDX gene expression profiling in women diagnosed with breast cancer in the United States?
 - d. What projections have been made in published analyses about the cost-effectiveness of using Mammprint and OncotypeDX gene expression profiling in women diagnosed with breast cancer?

[2] **Unclear/No abstract (promote to article review)**

[3] **NOT eligible (exclude): indicate reason for exclusion (below)**

Reason for EXCLUSION? (choose any that apply)

- [1] Study applies only to breast cancer biology
- [2] Study only applies to single or multiple gene predictors and does not involve OncotypeDX or Mammprint profiles
- [3] Does not involve OncotypeDX or Mammprint gene expression profiling tests
- [4] Does not involve original data or original data analysis
- [5] Does not involve women
- [6] Does not involve breast cancer patients
- [7] Not English language
- [8] Does not apply to the key questions
- [9] OTHER _____
- [10] Unclear

[4] **No, may be useful for BACKGROUND material (pull for hand searching If publish in 2002 or later)**

Appendix G: Data Abstraction Tables

Population Characteristics

| Study, Year | Intervention | General Characteristics | Diagnosis(es) | Treatments and Outcomes |
|-------------|--------------|-------------------------|---------------|-------------------------|
| | | | | |
| | | | | |
| | | | | |

Study Design

| Study, Year | Country | Study period (data collection period) | Study Type | Population size, N | Blinded (Y/N) | Study purpose |
|-------------|---------|---------------------------------------|------------|--------------------|---------------|---------------|
| | | | | | | |
| | | | | | | |
| | | | | | | |

Appendix G: Data Extraction Tables

Clinical Validity/Utility

| Study, year | Context | Methods | Results | Conclusions |
|-------------|---------|---------|---------|-------------|
| | | | | |
| | | | | |
| | | | | |

Analytic Validity

| Study, year | Measure | Conclusions |
|-------------|---------|-------------|
| | | |
| | | |
| | | |

Appendix G: Quality Assessment Matrix

| Section | Measure |
|-----------------------|---|
| Patients | <ul style="list-style-type: none"> • Describes population characteristics • Describes participant recruitment • Describes participant sampling • Describes inclusion/exclusion criteria • Describes treatments received • Describes randomization. |
| Materials and Methods | <ul style="list-style-type: none"> • Describes the reference standard. • Describes technical specifications of material and methods involved. • Describes type of biological material used (including control samples). • Includes definition of and rationale for the units, cutoffs and/or categories of the results of the index tests and the reference standard. • Describes blinding. • Describes methods for calculating or comparing measures. • Describes methods for calculating test reproducibility. • Describes methods of preservation and storage • Specifies the assay method used and provides (or references) a detailed protocol, including specific reagents or kits used, quality control procedures, reproducibility assessments, quantitation methods, and scoring and reporting protocols. |
| Results | <ul style="list-style-type: none"> • Describes the flow of patients through the study, including the number of patients included in each stage of the analysis(both overall and for each subgroup extensively examined). • Describes distributions of basic demographic characteristics (at least age and sex), standard (disease-specific) prognostic variables, and tumor marker, including numbers of missing values. • Presents univariate analyses showing the relation between the marker and outcome, with the estimated effect (e.g. hazard ratio and survival probability). • Provides similar analyses for all other variables being analyzed (for the effect of a tumor marker on a time-to-event outcome, a Kaplan–Meier plot is recommended). • For key multivariable analyses, report estimated effects (e.g. hazard ratio) with confidence intervals for the marker and, at least for the final model. • Provides estimated effects with confidence intervals from an analysis in which the marker and standard prognostic variables are included, regardless of their significance. |

Appendix H: Excluded Articles

- Blanchard A, Shiu R, Booth S et al. Gene expression profiling of early involuting mammary gland reveals novel genes potentially relevant to human breast cancer. *Front Biosci.* 2007;12:2221-32.
- Charafe-Jauffret E, Ginestier C, Monville F et al. Gene expression profiling of breast cell lines identifies potential new basal markers. *Oncogene.* 2006;25(15):2273-84.
- Dolled-Filhart M, Ryden L, Cregger M et al. Classification of breast cancer using genetic algorithms and tissue microarrays. *Clin Cancer Res.* 2006;12(21):6459-68.
- Drubin D, Smith JS, Liu W et al. Comparison of cryopreservation and standard needle biopsy for gene expression profiling of human breast cancer specimens. *Breast Cancer Res Treat.* 2005;90(1):93-6.
- Eden P, Ritz C, Rose C et al. "Good Old" clinical markers have similar power in breast cancer prognosis as microarray gene expression profilers. *Eur J Cancer.* 2004;40(12):1837-41.
- Ein-Dor L, Zuk O, and Domany E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc Nat. Acad Sci U.S.A.* 2006;103(15):5923-8.
- Espinosa E, Vara JA, Redondo A et al. Breast cancer prognosis determined by gene expression profiling: a quantitative reverse transcriptase polymerase chain reaction study. *J Clin Oncol.* 2005;23(29):7278-85.
- Fielden MR, Chen I, Chittim B et al. Examination of the estrogenicity of 2,4,6,2',6'-pentachlorobiphenyl (PCB 104), its hydroxylated metabolite 2,4,6,2',6'-pentachloro-4-biphenylol (HO-PCB 104), and a further chlorinated derivative, 2,4,6,2',4',6'-hexachlorobiphenyl (PCB 155). *Environ Health Perspect.* 1997;105(11):1238-48.
- Girault I, Lerebours F, Amarir S et al. Expression analysis of estrogen receptor (alpha) coregulators in breast carcinoma: Evidence that NCOR1 expression is predictive of the response to tamoxifen. *Clin Cancer Res* 2003;9(4):1259-1266.
- Gradishar WJ. Hormone therapy in postmenopausal women with breast cancer. *Adv. Stud. Med.* 2005;5(9 B):S817-S822.
- Jansen MP, Foekens JA, van Staveren IL et al. Molecular classification of tamoxifen-resistant breast carcinomas by gene expression profiling. *J Clin Oncol.* 2005;23(4):732-40.
- Kaklamani V. A genetic signature can predict prognosis and response to therapy in breast cancer: Oncotype DX. *Expert Rev Mol Diagn* 2006;6(6):803-9.
- Kominsky SL. Claudins: Emerging targets for cancer therapy. *Expert Rev. Mol. Med.* 2006;8(18):1-11.
- Kroll T, Odyvanova L, Clement JH et al. Molecular characterization of breast cancer cell lines by expression profiling. *J Cancer Res Clin Oncol.* 2002;128(3):125-134.
- Lipka C, Mankertz J, Fromm M et al. Impairment of the antiproliferative effect of glucocorticosteroids by 11(beta)-hydroxysteroid dehydrogenase type 2 overexpression in MCF-7 breast-cancer cells. *Horm Metabol. Res.* 2004;36(7):437-444.
- Master SR and Chodosh LA. Evolving views of involution. *Breast Cancer Res.* 2004;6(2):89-92.
- Mauriac L, Debled M, and MacGrogan G. When will more useful predictive factors be ready for use? *Breast* 2005;14(6):617-623.
- Molist R, Remvikos Y, Dutrillaux B et al. Characterization of a new cytogenetic subtype of ductal breast carcinomas. *Oncogene.* 2004;23(35):5986-5993.
- Naderi A, Teschendorff AE, Barbosa-Morais NL et al. A gene-expression signature to predict survival in breast cancer across independent data sets. *Oncogene.* Mar 1;26(10):1507-16.
- Nuyten DS, Kreike B, Hart AA et al. Predicting a local recurrence after breast-conserving therapy by gene expression profiling. *Breast Cancer Res.* 2006;8(5):R62.
- Oehr P. Proteomics as a tool for detection of nuclear matrix proteins and new biomarkers for screening of early tumors stage. *Anticancer Res.* 2003;23(2 A):805-812.
- Paik S, Kim CY, Song YK et al. Technology insight: Application of molecular techniques to formalin-fixed paraffin-embedded tissues from breast cancer. *Nat Clin Pract Oncol* 2005;2(5):246-54.
- Puzstai L. *Oncogenomics 2005 - Dissecting cancer through Genome Res.* : 2-6 February 2005, San Diego, CA, USA. *IDrugs* 2005;8(3):215-218.
- Ray ME, Yang ZQ, Albertson D et al. Genomic and Expression Analysis of the 8p11-12 Amplicon in human breast cancer cell lines. *Cancer Res.* 2004;64(1):40-47.

- Rha SY, Jeung HC, Yang WI et al. Alteration of hTERT full-length variant expression level showed different gene expression profiles and genomic copy number changes in breast cancer. *Oncol Rep.* 2006;15(4):749-755.
- Rodningen OK, Overgaard J, Alsner J et al. Microarray analysis of the transcriptional response to single or multiple doses of ionizing radiation in human subcutaneous fibroblasts. *Radiother Oncol.* 2005;77(3):231-40.
- Rody A, Karn T, Gatje R et al. Gene expression profiles of breast cancer obtained from core cut biopsies before neoadjuvant docetaxel, adriamycin, and cyclophosphamide chemotherapy correlate with routine prognostic markers and could be used to identify predictive signatures. *Zentralblatt für Gynäkologie* 2006;128(2):76-81.
- Ross DT and Perou CM. A comparison of gene expression signatures from breast tumors and breast tissue derived cell lines. *Dis Markers.* 2001;17(2):99-109.
- Rugo HS. Oncotype DX predicts tamoxifen benefits in ER+ breast cancer. *Comment. Oncol Rep.* 2005;-(FALL):18-19.
- Schneider J, Buness A, Huber W et al. Systematic analysis of T7 RNA polymerase based in vitro linear RNA amplification for use in microarray experiments. *BMC Genomics* 2004;5(1):29.
- Sengupta K, Banerjee S, Saxena NK et al. Differential expression of VEGF-A mRNA by 17(beta)-estradiol in breast tumor cells lacking classical ER-(alpha) may be mediated through a variant form of ER-(alpha). *Mol Cell Biochem.* 2004;262(1-2):215-224.
- Severgnini M, Biccato S, Mangano E et al. Strategies for comparing gene expression profiles from different microarray platforms: application to a case-control experiment. *Anal Biochem.* 2006;353(1):43-56.
- Sgroi DC, Haber DA, Ryan PD et al. RE: A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* 2004;6(5):445.
- Sun Y, Goodison S, Li J et al. Improved breast cancer prognosis through the combination of clinical and genetic markers. *Bioinformatics* 2007;23(1):30-7.
- Timar J, Ladanyi A, Petak I et al. Molecular pathology of tumor metastasis III. Target array and combinatorial therapies. *Pathology Oncology Research* 2003;9(1): 49-72.
- Verlinden I, Gungor N, Wouters K et al. Parity-induced changes in global gene expression in the human mammary gland. *Eur J Cancer Prev.* 2005;14(2):129-37.
- Weigelt B, Hu Z, He X et al. Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. *Cancer Res.* 2005;65(20):9155-8.
- Weil MR, Widlak P, Minna JD et al. Global survey of chromatin accessibility using DNA microarrays. *Genome Res.* 2004;14(7):1374-1381.
- Weisz A, Basile W, Scafoglio C et al. Molecular identification of ERalpha-positive breast cancer cells by the expression profile of an intrinsic set of estrogen regulated genes. *J Cell Physiol.* 2004;200(3):440-450.
- Welsh J, Wietzke JA, Zinser GM et al. Vitamin D-3 receptor as a target for breast cancer prevention. *J Nutr.* 2003;133(7 SUPPL.):2425S-2433S.
- West RB, Nuyten DS, Subramanian S et al. Determination of stromal signatures in breast carcinoma. *PLoS Biol* 2005;3(6):e187.
- Yared MA, Middleton LP, Bernstam FM et al. Expression of c-kit proto-oncogene product in breast tissue. *Breast J.* 2004;10(4):323-327.
- Zhang J-Y, Casiano CA, Peng X-X et al. Enhancement of antibody detection in cancer using panel of recombinant tumor-associated antigens. *Cancer Epidemiol Biomark Prev.* 2003;12(2):136-143.

Appendix I: Evidence Tables

Evidence Table 1. Study design, Oncotype DX™

| Study, Year | Country | Study period (data collection period) | Study Type | Population size, N | Blinded (Y/N) | Study purpose |
|--------------------------------|-------------|---------------------------------------|-------------------|---|---------------|-----------------------|
| Chang, 2007 ⁵⁵ | UK and USA | 2000-2004 | retrospective | 72 | NR | Validation |
| Cobleigh, 2005 ⁴⁷ | USA | 1979-1999 | retrospective | 78 | NR | Development |
| Cronin, 2004 ⁴⁴ | USA | NR | retrospective | 62 | NR | Protocols development |
| Cronin, 2007 ⁴⁵ | USA | NA | Analytic validity | NR | NR | Analytic validity |
| Esteva, 2005 ⁴⁸ | USA | 1978-1995 | retrospective | 149 | NR | Validation |
| Fan, 2006 ⁷⁹ | Netherlands | NR | retrospective | Poor prognosis, 180 Good prognosis, 115 | NA | Validation |
| Gianni, 2005 ⁴⁹ | Italy | 1998-2002 | prospective | INT-Milan, 95 | NA | Validation |
| Habel, 2006 ⁵⁰ | USA | 1985-1994 | case control | Cases, 220 Controls, 570 | NR | Validation |
| Hornberger, 2005 ⁶⁷ | USA | NR | retrospective | NSABP B-14, 668 NSABP B-20, 651 | NR | |
| Lyman, 2007 ⁷⁵ | USA | NR | retrospective | NSABP B-14, 668 NSABP B-20, 651 | NR | |
| Mina, 2006 ⁵¹ | NR | NR | retrospective | Gene profile, 45 Parent trial, 70 | NA | Validation |
| Oratz, in press ⁵⁶ | NR | NR | retrospective | 74 | NR | Validation |
| Paik, 2004 ^{*28} | USA | 1982-1988 | retrospective | 668 | NR | Validation |
| Paik, 2006 ^{*53} | USA | 1988-1993 | retrospective | Tamoxifen treated, 227 Tamoxifen and chemotherapy, 424 | NR | Validation |

* Supplemental information

NR = not reported; INT= Italian National Cancer Institute of Milan, Italy; NA = not applicable; NSABP = The National Surgical Adjuvant Breast and Bowel Project

Evidence Table 2. Study population characteristics, Oncotype DX™

| Study, Year | Intervention | General Characteristics | Diagnosis(es) | Treatments and Outcomes |
|------------------------------|--------------|--|---|--|
| Chang,2007 ⁵⁵ | | Age: mean 49 y Menopausal status: pre: n=26 (53.1%) post: n=23 (46.9) Race, n (%): W: 37 (52) B: 19 (27) H: 11 (16) A: 4 (6) UK: 1 Exclusion criteria: NR | TG: Well 2 (2.8%) TG: Moderate 29 (40.3%) TG: Poor 41 (56.9%) H: Ductal 67 (93.1%) H: Lobular 4 (5.6%) H: Mucinous 1 (1.4%) ER+: 47 (69.1%) ER-: 21 (30.9%) PR+: 38 (58.5%) PR-: 27 (41.5%) PR Unknown: 7 ; HER-2+: 7 (13.5%) HER-2-: 45 (86.5%) HER-2 Unknown: 20 Palpable nodes: Yes 7 (10.0%) Palpable nodes: No 63 (90.0%) Palpable nodes: Unknown 2 | Treatments: a: Docetaxel at 100 mg/m 128 2 was given every 3 weeks for a total of 4 cycles, and clinical response was assessed after the fourth cycle, at 12 weeks. s, a: Primary surgery and standard adjuvant therapy were then administered Outcome: CR (complete disappearance of the tumor), 12 (16.7%) PR (decrease in unidimensional size by at least 30%), 41 (56.9%) PD (an increase of more than 25%), 2 (2.8%) SD (any response that does not meet the definition of CR, PR, or PD), 17 (23.6%) Complete pathologic response 2 (3.2%) ; Others 61 (96.8%) Unknown 9 ; Relapse: Yes 13 (20.0%) ; No 52 (80.0%) Unknown 7 |
| Cobleigh, 2005 ⁴⁷ | | Age, mean y (range) SD: 57 (33-86) 13 Menopausal status: NR Race, n (%): NR Exclusion criteria: <ul style="list-style-type: none"> ● evidence of metastatic disease ● <10 positive nodes ● <5% of specimen shows tumor involvement ● no evidence of invasive breast cancer | TS 0-2; 26/78 (33%) TS 2-5; 28/78 (36%) TS >5; 24/78 (31%) Rush (B-R) TG 1; 22/78 (29%) Rush (B-R) TG 2; 28/78 (36%) Rush (B-R) TG 3; 28/78 (36%) Genomic Health (B-R) TG 1; 11/78 (14%) Genomic Health (B-R) TG 2; 37/78 (47%) Genomic Health (B-R) TG 3; 30/78 (38%) LN# 0-9; 0/78 (0%) LN# 10-15; 40/78 (51%) LN# 15-20; 18/78 (23%) LN# 20-30; 12/78 (15%) LN# >30; 8/78 (10%) | Treatments: h: tamoxifen Yes; 42/78 (54%) h: tamoxifen No; 36/78 (46%) a: chemotherapy (cyclophosphamide, methotrexate, fluorouracil) Yes; 62/78 (80%) a: chemotherapy No; 16/78 (20%) Outcomes: NR |

Evidence Table 2. Study population characteristics, Oncotype DX (continued)

| Study, Year | Intervention | General Characteristics | Diagnosis(es) | Treatments and Outcomes |
|----------------------------|--|---|--|-------------------------|
| Cronin, 2004 ⁴⁴ | There are no patient characteristics reported in this article. This is a study of analytic validity. | | | |
| Cronin, 2007 ⁴⁵ | There are no patient characteristics reported in this article. This is a study of analytic validity. | | | |
| Esteva, 2005 ⁴⁸ | None | Age, mean y, SD, median y: 58.0, 11.5, 59.0 Menopausal status: Pre; 122/149 (84.6%) Post; 27/149 (18.1%) Race, n (%): W, 126 (84.6) H, 8 (5.4) B 10 (6.7) A 5 (3.4) Exclusion criteria: • specimen shows <5% invasive tumor • insufficient RNA extracted from specimen • RT-PCR signal for genes outside the specifications for the assay • evidence of axillary lymph node involvement • use of adjuvant systemic therapy • follow up for <5 years | mean TS(cm) 2.3 median TS(cm) 2.0 standard deviation(cm) 1.1 B-R TG 1; 18/149 (12.1%) B-R TG 2; 86/149 (57.5%) B-R TG 3; 45/149 (30.2%) ER -; 46/149 (30.9%) ER +; 103/149 (69.1%) (HER2) -; 124/149 (83.2%) (HER2) +; 25/149 (16.8%) PR -; 99/149 (66.4%) PR +; 50/149 (33.6%) | NR |

Evidence Table 2. Study population characteristics, Oncotype DX. (continued)

| Study, Year | Intervention | General Characteristics | Diagnosis(es) | Treatments and Outcomes |
|-------------------------|----------------|--|--|--|
| Fan, 2006 ⁷⁹ | Poor prognosis | Age, n (%) <40yrs 52 (29) 40-44yrs, 41 (23) 45-49yrs, 55 (31) ≥50yrs; 32 (18) Menopausal status: NR Race: NR Exclusion criteria: • Previous history of cancer (except non-melanoma skin cancer) • Apical axillary lymph node involvement • Age at diagnosis >52 years • Tumor >5 cm in diameter at pathological examination | H I (good); 19/180 (11%) H II (intermediate); 56/180 (31%) H III (poor); 105/180 (58%) TS ≤20mm; 84/180 (47%) TS >20mm; 96/180 (53%) LN# 0; 91/180 (51%) LN#1-3; 63/180 (35%) LN# ≥4; 26/180 (14%) ER -; 66/180 (37%) ER +; 114/180 (63%) VI Absent; 108/180 (60%) VI 1-3 vessels; 18/180 (10%) VI >3 vessels; 54/180 (30%) Diagnostic method: RT-PCR | Treatment: h Yes; 23/180 (13%) h No; 157/180 (87%) a Yes; 66/180 (37%) a No; 114/180 (63%) s Breast conserving; 97/180 (54%) s mastectomy; 83/180 (46%) Outcome: NR |
| | Good prognosis | Age, n (%) <40yrs, 11 (10) 40-44yrs, 44 (38) 45-49yrs, 43 (37) ≥50yrs, 17 (15) Menopausal status: NR Race: NR Exclusion criteria: see poor prognosis | H I (good); 56/115 (49%) H II (intermediate); 45/115 (39%) H III (poor); 14/115 (12%) TS ≤20mm; 71/115 (62%) TS >20mm; 44/115 (38%) LN# 0; 60/115 (52%) LN# 1-3; 43/115 (37%) LN# ≥4; 12/115 (10%) ER-; 3/115 (3%) ER +; 112/115 (97%) VI Absent; 77/115 (67%) VI 1-3 vessels; 12/115 (10%) VI >3 vessels; 26/115 (23%) Diagnostic method: RT-PCR | Treatment: h: Yes; 17/115 (15%) h: No; 88/115 (85%) a: Yes; 44/115 (38%) a: No; 71/115 (62%) s: Breast conserving; 64/115 (56%) s: mastectomy; 51/115 (44%) Outcome: NR |

Evidence Table 2. Study population characteristics, Oncotype DX (continued).

| Study, Year | Intervention | General Characteristics | Diagnosis(es) | Treatments and Outcomes |
|----------------------------|--|--|--|--|
| Gianni, 2005 ⁴⁹ | <p>INT-Milan</p> <p>Doxorubicin (60mg/m²) + Paclitaxel (200mg/m²) every 3 weeks X3</p> <p>Followed by paclitaxel (80mg/m²) weekly X12</p> | <p>Age, mean y (range) SD: 49.9 (29-65) 9.0</p> <p>Menopausal status: NR</p> <p>Race, n (%): NR</p> <p>Exclusion criteria: NR</p> | <p>T1 0/89 (0%)</p> <p>T2 1/89 (1%)</p> <p>T3 2/89 (2%)</p> <p>T4a 0/89 (0%)</p> <p>T4b 70/89 (79%)</p> <p>T4d (inflammatory breast Ca) 16/89 (18%)</p> <p>LNO 14/89 (16%)</p> <p>LN1 51/89 (57%)</p> <p>LN2 24/89 (27 %)</p> <p>Supraclavicular LN 14/89 (16%)</p> <p>TG 1; 21/80 (24%)</p> <p>TG 2; 49/89 (55%)</p> <p>TG 3; 19/89 (21%)</p> <p>ER -; 31/89 (35%)</p> <p>ER +; 52/89 (58%)</p> <p>ER Not Assessable 6/89 (7%)</p> | <p>Treatments:</p> <p>s: Mastectomy; 79/89 (89%)</p> <p>s: Quadrantectomy; 8/89 (9%)</p> <p>No surgery; 2/89 (2%)</p> <p>Outcomes:</p> <p>CR; 17/89 (19%)</p> <p>PR; 57/89 (64%)</p> <p>SD; 13/89 (15%)</p> <p>PD; 2/89 (2%)</p> <p>pCR; 11/89 (12%)</p> |
| | <p>MDACC-Houston</p> <p>12 weeks of paclitaxel</p> <p>followed by fluorouracil, doxorubicin and cyclophosphamide (FAC) X4 courses</p> | <p>Age, mean y (range) SD: 53.1 (29-79) 11.7</p> <p>Menopausal status: NR</p> <p>Race, n (%): NR</p> <p>Exclusion criteria: NR</p> | <p>T1 7/82 (9%)</p> <p>T2 46/82 (56%)</p> <p>T3 15/82 (18%)</p> <p>T4a 1/82 (1%)</p> <p>T4b 6/82 (7%)</p> <p>T4d (inflammatory breast Ca) 7/82 (9%)</p> <p>Nodal status</p> <p>N0 28/82 (34%)</p> <p>N1 38/82 (46%)</p> <p>N2 8/82 (10 %)</p> <p>Supraclavicular LN 8/82 (10%)</p> <p>TG 1; 2/82 (3%)</p> <p>TG 2; 31/82 (37%)</p> <p>TG 3; 50/82 (60%)</p> <p>ER -; 48/82 (58%)</p> <p>ER +; 32/82 (39%)</p> <p>ER Not Assessable 3/82 (4%)</p> | <p>Treatments:</p> <p>s: Mastectomy; 48/89 (59%)</p> <p>s: Quadrantectomy; 33/89 (40%)</p> <p>No surgery; 1/89 (1%)</p> <p>Outcomes:</p> <p>CR; 28/82 (34%)</p> <p>PR; 40/82 (49%)</p> <p>SD; 13/82 (16%)</p> <p>PD; 1/82 (1%)</p> <p>pCR; 21/82 (26%)</p> |

Evidence Table 2. Study population characteristics, Oncotype DX (continued).

| Study, Year | Intervention | General Characteristics | Diagnosis(es) | Treatments and Outcomes |
|--------------------------------|--|--|---|--|
| Habel, 2006 ⁵⁰ | Cases | <p>Age, n (%):</p> <ul style="list-style-type: none"> <40, 17 (8) 40-49, 42 (19) 50-59, 64 (29) 60-74, 97 (44) <p>Menopausal status: NR</p> <p>Race, n (%)</p> <ul style="list-style-type: none"> W, 172 (78) H, 7 (3) B, 20 (9) A, 21 (10) <p>Exclusion criteria:</p> <ul style="list-style-type: none"> • positive node status • diagnosed at or after 75 years of age • initial disease treated with chemotherapy • inflammatory or bilateral breast cancer or evidence of metastases • prior invasive cancer • unknown/unconfirmed treatment with tamoxifen | <p>TS < 1 cm, 49 (22)</p> <p>TS 1.1-2 cm, 93 (42)</p> <p>TS 2.1-4 cm, 72 (33)</p> <p>TS > 4 cm, 6 (3)</p> <p>TG well diff, 25 (11)</p> <p>TG mod diff, 92 (42)</p> <p>ER+, 168 (76)</p> <p>ER-, 52 (24)</p> <p>Poor diff, 103 (47)</p> | <p>Treatments:</p> <p>h: Tamoxifen, 64 (26)</p> <p>Outcomes:</p> <p>RS <18, 57 (26%)</p> <p>RS 18-3, 54 (25%)</p> <p>RS > 31, 109 (50%)</p> |
| | Controls | <p>Age, n (%):</p> <ul style="list-style-type: none"> <40, 23 (4) 40-49, 127 (22) 50-59, 152 (27) 60-74, 268 (47) <p>Menopausal status: NR</p> <p>Race, n (%)</p> <ul style="list-style-type: none"> W, 457 (810) H, 13 (2) B, 47 (8) A, 53 (9) <p>Exclusion criteria: see cases</p> | <p>TS < 1 cm, 193 (34)</p> <p>TS 1.1-2 cm, 255 (45)</p> <p>TS 2.1-4 cm, 114 (20)</p> <p>TS > 4 cm, 8 (1)</p> <p>TG well diff, 175 (31)</p> <p>TG mod diff, 261 (46)</p> <p>ER+, 514 (90)</p> <p>ER-, 56 (10)</p> <p>Poor diff, 134 (23)</p> | <p>Treatments:</p> <p>h: Tamoxifen, 177 (31)</p> <p>Outcomes:</p> <p>RS <18, 322 (56%)</p> <p>RS 18-3, 108 (19%)</p> <p>RS > 31, 140 (25%)</p> |
| Hornberger, 2005 ⁶⁷ | There are no patient characteristics reported in this article. This was a study on the economics of using a recurrence score to predict distant recurrence-free survival in patients with early-stage breast cancer. | | | |

Evidence Table 2. Study population characteristics, Oncotype DX (continued).

| Study, Year | Intervention | General Characteristics | Diagnosis(es) | Treatments and Outcomes |
|-------------------------------|-----------------------------------|---|---|---|
| Lyman, 2007 ⁷⁵ | High RS (≥ 31) | Age: NR Menopausal status: NR Race: NR Exclusion criteria: NR | DR: Tamoxifen alone, 38.3%; Tamoxifen + chemo, 11.1% | NR |
| | Intermediate RS (18-30) | | DR: Tamoxifen alone, 17.8%; Tamoxifen + chemo, 10.1% | |
| | Low RS (<18) | | DR: Tamoxifen alone, 3.7%; Tamoxifen + chemo, 5% | |
| Mina, 2006 ^{51†} | Parent trial (Miller, 2005), n=70 | Age, mean y (range): 50 (30-65) Menopausal status: NR Race/ethnicity, n (%): NR Exclusion criteria: NA | TS: 6 cm TG: NR ER+: 40 (57) HER2+: 14 (20) pCR: 9 (13) Pathologically negative nodes: 37 (53) Inflammatory disease 14 (20) Palpable lymph nodes: 33 (47) | NA |
| | Gene Profile subset, n=45 | Age, mean y (range): 49 (29-64) Menopausal status: NR Race/ethnicity, n (%): NR Exclusion criteria: NR | TS: 6.8 cm (2.3-21) TG: well differentiated 11 (24); mod differentiated 12 (27); poor differentiated 22 (49) ER+: 26 (57) HER2+: 8 (18) pCR: 6 (13) Pathologically negative nodes: 23 (51) Inflammatory disease 9 (20) Palpable lymph nodes: 21 (47) | Treatments: a: sequential doxorubicin a:, 75 mg/M2 ea 2weeks for 3 cycles; docetaxel (T) 40 mg/M2 1weeks for 6 cycles a: definitive surgery was at the completion of all chemotherapy r: at the discretion of the treating oncologist h: Tamoxifen was recommended for all ER+ pts Outcomes' |
| Oratz, in press ⁵⁶ | | Median age (range): 54 (35-77) Menopausal status: NR Race, n (%): NR Exclusion criteria: NR | TS (cm) 0-1.0; 32/68 (47%) TS (cm) 1.1-2.0; 28/68 (41%) TS (cm) 2.1-3.0; 8/68 (12%) TG 1; 30/68 (44%) TG 2; 24/68 (35%) TG 3; 14/68 (21%) LN+/LN--: 0/74 ER+/ER--:74/0 | NR |

Evidence Table 2. Study population characteristics, Oncotype DX (continued).

| Study, Year | Intervention | General Characteristics | Diagnosis(es) | Treatments and Outcomes |
|--------------------------|---|--|---|--------------------------------|
| Paik, 2004 ²⁸ | NSABP B-14, Genomic Health populations: n = 668 | Age, n (%): <50y, 194 (29) 50 to 60y, 173 (26) ≥60y, 301 (45) Menopausal status: NR Race/ethnicity, n (%): W, 615 (92) B, 31 (5) O, 22 (3) Exclusion criteria: • insufficient tumor as assessed by histopathology <5% overall tissue • insufficient RNA < 0.5µg • weak TR-PCR signal (avg cycle threshold for the reference genes >35) | TS 0-1.0; 109 (16%) TS 1.1-2.0; 305 (46%) TS 2.1-4.0; 220 (33%) TS ≥4.1; 34 (5%) | NR |
| | NSABP B-14, remaining population, n = 1949 | Age, n (%): <50y, 618 (32) 50 to 60y, 583 (30) ≥60y, 748 (38) Menopausal status: NR Race/ethnicity, n (%): W, 1763 (90) B, 87 (5) O, 99 (5) Exclusion criteria: NR | TS 0-1.0; 366 (19%) TS 1.1-2.0; 877 (45%) TS 2.1-4.0; 635 (32%) TS ≥4.1; 71 (4%) | NR |
| | All B-14 patients, n = 2617 | Age, n (%): <50y, 812 (31) 50 to 60y, 756 (29) ≥60y, 1049 (40) Menopausal status: NR Race/ethnicity, n (%): W, 2378 (91) B, 118 (4) O, 121 (5) Exclusion criteria: NR | TS 0-1.0; 475 (18%) TS 1.1-2.0; 1182 (45%) TS 2.1-4.0; 855 (33%) TS ≥4.1; 105 (4%) | |

Evidence Table 2. Study population characteristics, Oncotype DX (continued)

| Study, Year | Intervention | General Characteristics | Diagnosis(es) | Treatments and Outcomes |
|--------------------------|--|---|--|--|
| Paik, 2006 ⁵³ | TAM treated and chemotherapy treated NSABP-B20 patients, n=651 | Age, n (%): <40, 63 (10) 40-49, 226 (35) 50-59, 166 (25) >60, 196 (30) Menopausal status: NR Race/ethnicity, n (%): NR Exclusion criteria: <ul style="list-style-type: none"> insufficient tumor as assessed by histopathology <5% overall tissue insufficient RNA < 0.5µg weak TR-PCR signal (avg cycle threshold for the reference genes >35) | TS 0-1 cm, 110 (17) TS 1.1-2 cm, 318 (49) TS 2.1-4 cm, 196 (30) TS >4 cm, 24 (4) TG well diff, 77 (13) TG mod diff, 339 (59) TG poor diff, 163 (28) ER 10-49 fmol/mg, 261 (40) ER 50-99 fmol/mg, 153 (23) ER 100-199 fmol/mg, 122 (19) ER >200 fmol/mg, 115 (18) PR 0-9 fmol/mg, 105 (16) PR 10-49 fmol/mg, 113 (17) PR 50-99 fmol/mg, 97 (15) PR 100-199 fmol/mg, 110 (17) PR >200 fmol/mg, 226 (35) | Treatments: NA Outcomes: RS <18, 353 RS 18-3, 134 RS > 31, 164 |
| | TAM treated, n=227 | Age, mean y (range): NR Menopausal status: NR Race/ethnicity, n (%): NR Exclusion criteria: see all patients | NR | Treatments: h: Tamoxifen Outcomes: RS <18, 135; RFS 10 year-96.8% RS 18-3, 45; RFS 10 year-90.9% RS > 31, 47; RFS 10 year-60.5% |
| | TAM and chemotherapy, n=424 | Age, mean y (range): NR Menopausal status: NR Race/ethnicity, n (%): NR Exclusion criteria: see all patients | NR | Treatments: h: Tamoxifen a: cyclophosphamide, methotrexate, and fluorouracil (CMF) OR methotrexate and fluorouracil (MF) Outcomes: RS <18, 218; RFS 10 year-95.6% RS 18-3, 89; RFS 10 year-89.1% RS > 31, 117; RFS 10 year-88.1% |

* Diagnoses: H = histology/histologic type; TS = tumor size/diameter; LN# = # of Lymph nodes at primary diagnosis; LN = lymph node status; TG = tumor grade; ER = estrogen receptor status; HER = HER neu status; PR = Progesterone receptor status; SP = S-phase fraction; Pl = ploidy; VI = vascular invasion; PTS = site of primary tumor; HR = hormonal status, ER or PR or both; DR = distant recurrence; LI lymphatic infiltrate

† Treatments: H = hormonal; A = adjuvant; S = surgical; R = radiation

* Treatment outcomes: OS = overall survival; RFS = relapse-free survival (first disease relapse, irrespective of site); CR = complete response; PR = partial response; RS: recurrence score; SD = stable disease; PD = progressive disease; pCR = complete pathologic response; D = death

W=white; B=black; H=Hispanic; A=Asian; NR=not reported; RNA=ribonucleic acid; NSABP= National Surgical Adjuvant Breast and Bowel Project; CMF= cyclophosphamide, methotrexate, and fluorouracil; MF= methotrexate and fluorouracil; TAM=tamoxifen

Evidence Table 3. Analytic validity, Oncotype DX™

| Study, year | Measure | Conclusions |
|---------------------------|--|--|
| Chang, 2007 ⁵⁵ | <p>Context: This study used the Oncotype DX assay analyzing FFPE specimens. Out of the 97 eligible samples the analysis was carried on 80 samples*</p> <p>Methods:</p> <ul style="list-style-type: none"> ● Standard Oncotype DX assay protocols were developed ● Methods for IHC not described ● The degree of concordance between IHC and RT-PCR with respect to ER, PR, and HER-2 was assessed graphically and by the kappa statistic ● Positivity from RT-PCR analysis was assigned by the following expression levels: <ul style="list-style-type: none"> ER, >6.5.0 CT PR, > 5.5 CT HER2, > 11.5 CT <p>Results:</p> <ul style="list-style-type: none"> ● 80 patients out of 97 successfully analyzed, for the following reasons*: <ul style="list-style-type: none"> Tumor content < 5%: 16 /97 patients (16.5%) Poor RNA yield: 1/80 patient (1.2%) Successful assays: 80 /97 (82.4%) ● Agreement between IHC vs RT-PCR: <ul style="list-style-type: none"> ER, k = 1.00, (95% CI = 1.00 to 1.00) PR, k= 0.57, (95% CI = 0.37 to 0.77) HER2, k= 0.74, (95% CI = 0.45 to 1.00) <p>*Of the 80 patients that were assayed however clinical information was available for 72 cases</p> | The RT-PCR technology provides a potential platform for a predictive test using small amounts of routinely processed specimens (core biopsies) |

Evidence Table 3. Analytic validity, Oncotype DX (continued)

| Study, year | Measure | Conclusions |
|------------------------------|--|---|
| Cobleigh, 2005 ⁴⁷ | <p>Context: This study used the Oncotype DX assay analyzing FFPE specimens. Out of the 86 eligible samples the analysis was carried on 78 samples.</p> <p>Methods:</p> <ul style="list-style-type: none"> ● Standard Oncotype DX assay protocols were used ● Immunohistochemistry (IHC) of ER, PR, HER2/neu and Ki-67 by standard biotin-streptavidin method with the appropriate antibodies (DAKO, CA, USA): ● The ER staining were defined as positive if staining of any intensity was seen in more than 10% of cells ● ER positivity from RT-PCR analysis was assigned to expression levels > 6.5 CT ● Pearson's correlation for co-expressed genes evaluation <p>Results:</p> <ul style="list-style-type: none"> ● 78 patients out of 86 successfully analyzed, for the following reasons: <ul style="list-style-type: none"> ○ Clinically ineligible: 1 stage VI; ○ Tumor content < 5%: 7 patients ○ Successful assays: 78/85 (91.7%) ● Correlation between two distinct probe-prime sets for ER in RT-PCR: $r = 0.96$ ● Agreement between IHC vs RT-PCR: <ul style="list-style-type: none"> ○ ER, $k = 0.83$ ○ PR, $k = 0.40$ ○ HER2, $= 0.67$ ○ Ki-67/MIB1, $= 0.22$ ● Pearson's correlation for gene pairs: <ul style="list-style-type: none"> ○ KRT5 and KRT17 ($r = 0.86$) ○ LPL and RBP4 ($r = 0.82$) ○ HER2 and GRB7 ($r = 0.82$); | <p>The accuracy and specificity of this RT-PCR assay of formalin-fixed, paraffin embedded tumor tissue was supported by comparison of the results of RT-PCR assay of RNA and IHC assay of protein for ER, PR, and HER2;</p> |

Evidence Table 3. Analytic validity, Onco^{type} DX (continued)

| Study, year | Measure | Conclusions |
|----------------------------|--|-------------|
| Cronin, 2004 ⁴⁴ | <p>Context: This study used the Onco^{type} DX assay analyzing FFPE specimens.</p> <p>Methods:</p> <ul style="list-style-type: none"> • Standard Onco^{type} DX assay protocols were developed • Methods for IHC and FISH not described • ER positivity from RT-PCR analysis was assigned to expression levels > 8 CT • Pearson's correlation for co-expressed genes evaluation and correlation between FFPE and frozen specimen <p>Results:</p> <ul style="list-style-type: none"> • ER status by RT-PCR: <ul style="list-style-type: none"> ER-positive: 79% (49 out of 62 tumors) ER-negative: 21% (13 out of 62 tumors) • ER status, IHC vs RT-PCR comparison: <ul style="list-style-type: none"> 58 concordant samples 24ER positive by RT-PCR were negative by IHC • Agreement between IHC vs RT-PCR: <ul style="list-style-type: none"> ER, 93.5% PR, 84% HER2, 100% • Pearson's correlation for gene pairs: <ul style="list-style-type: none"> KRT5 and KRT17 (r = 0.85) LPL and RBP4 (r = 0.84) HER2 and GRB7 (r = 0.71) ER1 and GATA23(r = 0.60) • Comparison between FFPE and frozen specimens, 48 genes: <ul style="list-style-type: none"> Correlation = 0.91, p value < 0.0001) | |

Evidence Table 3. Analytic validity, Onco^{type} DX (continued)

| Study, year | Measure | Conclusions |
|----------------------------|--|---|
| Cronin, 2007 ⁴⁵ | <p>Context: This study used the Onco^{type} DX assay</p> <ul style="list-style-type: none"> • The following analytic components of the assay were addressed: <ul style="list-style-type: none"> Limits of detection (LOD) Limits of quantification (LOQ) Amplification efficiency Dynamic range linearity Accuracy and precision Reproducibility <p>Methods:</p> <ul style="list-style-type: none"> • Repeated negative sample measurement and a serial dilutions experiment, using a pooled reference RNA were performed for each gene, to assess LOD, LOQ, amplification efficiency, dynamic range linearity, and accuracy and precision • Repeated measurements of 2 aliquots of a single RNA sample across multiple days, operators, RT-PCR plates, 7900HT instruments, and liquid-handling robots were performed to assess individual gene and RS reproducibility • In the linearity study cDNA input varied over RNA equivalents between 2e-10 and 2000 ng per reaction; in all other experiments cDNA input was equivalent to 2 ng RNA • The linearity response was assessed for each individual gene in the assay, by examining the SE from nonlinear polynomial regression • ANOVA was used to asses precision variability due to plate and well within plate, treating plates and wells within plates at random • Mixed-effect ANOVA was used to estimate components of variance due to days, plates within days, and within plates by treating operator, 7900HT instrument, and liquid-handling robot as fixed effects <p>Results:</p> <ul style="list-style-type: none"> • The LOD and the LOQ for all the genes proved to be well above the pre-specified limits of CT units (40 cycles): <ul style="list-style-type: none"> LOD ranged from 38.7 CT (RPLPO) to 40.0 CT (most of the genes in the assay) LOQ ranged from 36.9 CT (ERBB2) to 39.5 CT (GAPDH) • Amplification efficiencies for all the genes ranged from 75.3% (GAPDH) to 112.1% (BAG1), with values exceeding 100% due to the cumulative nature of the analysis along the sample-dilution series | <p>Authors reported that the following procedures a currently performed to assure the reproducibility of the assay:</p> <ul style="list-style-type: none"> A standard RNA control sample is assayed at least once per batch of patient (46 samples) PCR controls are run in every assay plate RT-PCR failures are excluded from analysis Expression values are assigned when at least 2 of 3 assay wells provide acceptable RT-PCR results <p>All 21 genes must have an expression value assigned for an RS to be calculated and reported</p> |

Evidence Table 3. Analytic validity, Oncotype DX (continued)

| Study, year | Measure | Conclusions |
|-------------------------|---|-------------|
| Cronin, 2007 (con't) | <ul style="list-style-type: none"> ● Assay linearity and dynamic range: <ul style="list-style-type: none"> 6 genes (ACTB, BAG1, CD68, ESR1, RPLP0, and TFRC), were linear over the entire RNA range 6 genes (BCL2, SCUBE2, GRB7, ERBB2, MMP11, and GAPDH), were linear over the entire RNA range tested, with an estimated maximal deviation from linearity below 1 CT over a linear range > 2000-fold, as specified by CLSI 4 genes (MYBL2, AURKA, BIRC5, MYBL2) were linear over a more restricted RNA concentration range of 2e-8 to 2000 ng, with an estimated maximal deviation from linearity below 1 CT over a linear range > 2000-fold, as specified by CLSI 1 gene (PR) was linear over a more restricted RNA concentration range of 2e-9 to 2000 ng 2 genes (CCNB1, GSTM1) were linear over a more restricted RNA concentration range of 2e-8 to 2000 ng 2 genes (CTSL2, GUSB) were linear over a more restricted RNA concentration range of 2e-7 to 2000 ng ● Assay quantitative bias and precision at the 2-ng/well for the 16 cancer-related genes: <ul style="list-style-type: none"> range = -10% (BAG1) to 6% (CTSL2) estimated mean deviation = 0.3% CV of 5.7% ● Assay quantitative bias and precision at the 2-ng/well for the 5 reference genes: <ul style="list-style-type: none"> range = -1.5% (GUSB) to 3.3% (ACTB) estimated mean deviation = 0.7% CV of 3.2% ● Reproducibility, CT measurements SD for the individual genes: <ul style="list-style-type: none"> Total SD ranged from 0.06 to 0.15 CT units Between day SD ranged from 0 to 0.055 CT units Between plated SD ranged from 0 to 0.090 CT units Within plated SD ranged from 0.057 to 0.147 CT units At a CT of 30 a maximum SD of 0.15 translate into a CV of 0.5% The largest differences between operators, liquid handling robots, and 7900HT instruments, were < 0.5 CT ● Reproducibility, CT measurements SD for the RS: <ul style="list-style-type: none"> Total SD was 0.792 RS units Between day SD was 0 RS units Between plated was 0 RS units ● Within plated was 0.792 RS units | |

Evidence Table 3. Analytic validity, Oncotype DX (continued)

| Study, year | Measure | Conclusions |
|----------------------------|--|-------------|
| Esteva, 2005 ⁴⁸ | <p>Context: This study used the Oncotype DX assay analyzing FFPE specimens. Out of the 220 eligible samples the analysis was carried on 149 samples</p> <p>Methods:</p> <ul style="list-style-type: none"> • Standard Oncotype DX assay protocols were used • IHC methods in Esteva et al., 2003⁴⁸ and Wang et al., 2002⁸⁹ • Positivity from RT-PCR analysis was assigned by the following expression levels: <ul style="list-style-type: none"> ER, >7.0 CT PR, > 6.0 CT HER2, > 11.5 CT • Concordance between RT-PCR and IHC results by Cohen's k statistics • Logistic model for HER2 measurement by RT-PCR, IHC as a quantal response • RT-PCR specificity and sensitivity, in comparison to IHC for HER2) were obtained at the different thresholds <p>Results:</p> <ul style="list-style-type: none"> • 149 patients out of 220 successfully analyzed, for the following reasons: <ul style="list-style-type: none"> Tumor content < 5%: 42 /220 patients (19.0%)* Poor RNA yield: 4/220 patients (1.8%) RT-PCR not within the specifications; 3 /220(1.3%) Low reference genes RT-PCR signal: 22/220 (10%) • Successful assays: 149 /220 (67.7%) • Agreement between IHC vs RT-PCR: <ul style="list-style-type: none"> ER, k = 0.80, (=+ 0.05) PR, k = 0.48 HER2, k = 0.6, (=+ 0.08) • Agreement between IHC vs RT-PCR: <ul style="list-style-type: none"> Logistic model p value: < 0.001 • HER2, RT-PCR > 11.50CT: <ul style="list-style-type: none"> Specificity: 77% Sensitivity: 84% • HER2, RT-PCR > 11.5 CT: <ul style="list-style-type: none"> Specificity: 89% Sensitivity: 84% • HER2, RT-PCR > 12.0 CT: <ul style="list-style-type: none"> Specificity: 95% Sensitivity: 68% | |

Evidence Table 3. Analytic validity, Oncotype DX (continued)

| Study, year | Measure | Conclusions |
|----------------------------|--|-------------|
| Gianni, 2005 ⁴⁹ | <p>Context: This study used the Oncotype DX assay analyzing FFPE specimens. Out of the 95 eligible samples the analysis was carried on 89 samples.</p> <p>Methods:</p> <ul style="list-style-type: none"> ● Standard Oncotype DX assay protocols were used; ● Immunohistochemistry (IHC) of ER, PR by reagents from Lab Vision-Neomarkers (Fremont, CA) ● The ER staining were defined as positive if staining of any intensity was seen in more than 10% of cells; ● ER positivity from RT-PCR analysis was assigned to expression levels > 6.5 CT; <p>Results:</p> <ul style="list-style-type: none"> ● 89 patients out of 95 successfully analyzed, for the following reasons: <ul style="list-style-type: none"> ○ Poor RNA yield: 4 patients (4.2%); ○ Tumor content < 5%: 2 patients (2.1%); ○ Successful assays: 89/95 (93.7%); ● Agreement between IHC vs RT-PCR: <ul style="list-style-type: none"> ○ ER, k = 0.84, (95% CI = 0.71 to 0.96); ○ PR, k = 0.71, (95% CI = 0.56 to 0.86); | |

Evidence Table 3. Analytic validity, Oncotype DX (continued)

| Study, year | Measure | Conclusions |
|---------------------------|--|-------------|
| Habel, 2006 ⁵⁰ | <p>Context: This study used the Oncotype DX assay, analyzing FFPE specimens; data about the success rate of the assay were reported.</p> <p>Methods:</p> <ul style="list-style-type: none"> • Standard Oncotype DX assay protocols were used • Pearson’s correlation and ANOVA to assess within patients correlation and variability: <ul style="list-style-type: none"> 60 blocks that did not undergo macro-dissection from 20 (2 to 5 blocks per patient) 49 core biopsies or tumor resection blocks • ER positivity from RT-PCR analysis was assigned to expression levels > 6.5 CT <p>Results:</p> <ul style="list-style-type: none"> • After medical history review, FFPE blocks were available for 234 cases and 631 controls, for a total of 865 patients, who underwent pathological review: <ul style="list-style-type: none"> 59 patients needed macro-dissection 19 patients showed tumor content < 5% and were excluded from the study • The final analysis considered 220 cases and 570 controls, for a total 790 patients, resulting form a 7.9% loss due to insufficient tumor content and 1% failure rate of the RT-PCR assay; it is not clear which is the total with respect these percentage are expressed, since it was not reported • RS (as a continuous value) standard deviation (SD) and Pearson’s correlation observed in two unpublished study: <ul style="list-style-type: none"> Overall between blocks SD was 3.0 RS units For 16 of the 20 patients, the between blocks SD was < 2.5 RS units Pearson’s correlation = 0.86 Similar results from the second study • The concordance of ER status between RT-PCR and medical chart information: <ul style="list-style-type: none"> kappa = 0.49, 95% CI 0.41–0.56 115/122 of the discordances were patients classified as ER+ based on RT-PCR | |

Evidence Table 3. Analytic validity, Oncotype DX (continued)

| Study, year | Measure | Conclusions |
|--------------------------|---|--|
| Mina, 2006 ⁵¹ | <p>Context: This study used the Oncotype DX assay, analyzing FFPE specimens. Out of the 70 eligible samples the analysis was carried on 45 samples.</p> <p>Methods:</p> <ul style="list-style-type: none"> • Standard Oncotype DX assay protocols were used • Immunohistochemistry (IHC) of ER, PR, HER2/neu and Ki-67 by standard biotin-streptavidin method with the appropriate antibodies (DAKO, CA, USA): <ul style="list-style-type: none"> anti-ER: mouse monoclonal 1 D5; anti-PR: mouse monoclonal PR636 anti-Ki-67: mouse monoclonal MIB-1; HER2/neu: HercepTest; • The ER staining were defined as positive if staining of any intensity was seen in more than 10% of cells • ER positivity from RT-PCR analysis was assigned to expression levels > 6.5 CT <p>Results:</p> <ul style="list-style-type: none"> • 45 patients out of 70 successfully analyzed, for the following reasons: <ul style="list-style-type: none"> No consent: 3 patients No specimen available: 10 patients Poor RNA yield: 9/57 patients (15.8%) Tumor content < 20%: 3/57 patients (5.3%) Successful assays: 45/57 patients (78.9%) • Reference genes showed smaller ranges of expression across samples, than the 16 cancer-related genes, especially the ER-related group (standard deviation in CT is reported: <ul style="list-style-type: none"> Normalization genes: b-actin = 0.56, GAPDH = 0.71, GUS = 0.75, RPLPO = 0.53, and TFRC = 0.64 ER-related genes: ER = 2.72, PR = 2.27, Bcl2 = 1.41, and SCUBE2 = 2.45 • ER status by RT-PCR: <ul style="list-style-type: none"> ER-positive: 64% (29 out of 45 tumors) ER-negative: 36% (16 out of 45 tumors) • ER status, IHC vs RT-PCR comparison: <ul style="list-style-type: none"> 41 concordant samples 2 ER positive by IHC were negative by RT-PCR 2 ER positive by RT-PCR were negative by IHC • Agreement data for PR, KI-67 and HER2/neu were not reported | Gene expression analysis on core biopsy samples is feasible; |

Evidence Table 3. Analytic validity, Oncotype DX (continued)

| Study, year | Measure | Conclusions |
|-------------------------------|---|-------------|
| Oratz, in press ⁵⁶ | <p>Context: This study used the Oncotype DX assay analyzing FFPE specimens. Out of the 74 eligible samples the analysis was carried on 72 samples</p> <p>Methods:</p> <ul style="list-style-type: none"> • Standard Oncotype DX assay protocols were developed <p>Results:</p> <ul style="list-style-type: none"> • 72 patients out of 74 successfully analyzed, for the following reasons: <ul style="list-style-type: none"> Tumor content < 5%: 2 /74 patients (2.7%) Successful assays: 72 /74 (97.3%) | |
| Paik, 2004 ^{*28} | <p>Context: This study used the Oncotype DX assay analyzing FFPE specimens. FFPE specimens were used. Out of the 754* eligible samples the analysis was carried on 668 samples</p> <p>Methods:</p> <ul style="list-style-type: none"> • Standard Oncotype DX assay protocols were used • ER and PR receptors proteins were measured by ligand-binding assays • HER2 DNA was measured by a fluorescence in situ hybridization assay (PathVysion, Vysis) • Reproducibility within and between blocks was assessed by performing the 21-gene assay in five serial sections from six blocks in two patients <p>Results:</p> <ul style="list-style-type: none"> • 675 patients out of 754 successfully analyzed, for the following reasons: <ul style="list-style-type: none"> Tumor content < 5%: 79 /754 patients (89.5%)* Successful assays: 668 /675 (98.9%) • Reproducibility evaluation: <ul style="list-style-type: none"> 16 Cancer genes SD ranged from 0.07 to 0.21 CT units Within-block RS SD = 0.72 RS unit (95% CI = 0.55 to 1.04) Total within-patient SD (including between and within-block SD) = 2.2 RS units Similar variability in the RS was observed in reanalysis of clinical trial samples on separate days with different reagent lots (data not shown). <p>* The total of 754 (79+675) patients was computed and not clearly reported in the study</p> | |

Evidence Table 3. Analytic validity, *Oncotype DX* (continued)

| Study, year | Measure | Conclusions |
|---------------------------|---|-------------|
| Paik, 2006 ^{*53} | <p>Context: This study used the <i>Oncotype DX</i> assay analyzing FFPE specimens. Out of the 670 eligible samples the analysis was carried on 651 samples</p> <p>Methods: Standard <i>Oncotype DX</i> assay protocols were used.</p> <p>Results:</p> <ul style="list-style-type: none"> • A successful assay was obtained for 651 patients out of 670 for which a specimen block was available: Successful assays: 651/670 (97.2%) | |

FFPE = formalin fixed paraffin embedded; IHC = immunohistochemistry; RT-PCR = real time polymerase chain reaction; ER = estrogen receptor; PR = progesterone receptor; HER-2 = human epidermal growth factor receptor 2; CT = cycle threshold; RNA= ribonucleic acid; CI=confidence interval; FISH = Fluorescence In Situ Hybridization; LOD=limits of detection; LOQ=limits of quantification; ANOVA= analysis of variance; CLSI= Clinical and Laboratory Standards Institute; SD = standard deviation; RS = recurrence score; CV= coefficients of variation;

Evidence Table 4. Clinical utility and validity, Oncotype DX™

| Study, year | Measure | Conclusions |
|------------------|--|---|
| Chang,2007 55 | <p>Context: This study used the <i>Oncotype DX</i> assay and assessed its predictive utility</p> <p>End points: prediction of clinical response (by the RECIST method) to docetaxel treatment in women with breast cancer</p> <p>Methods:</p> <ul style="list-style-type: none"> ● The degree of concordance between IHC and RT-PCR with respect to ER, PR, and HER-2 was assessed graphically and by the kappa statistic; ● Logistic regression was used to evaluate the association between clinical complete response with: <ul style="list-style-type: none"> Various clinical variables; The expression levels of individual genes; The average expression levels for <i>Oncotype DX</i> gene groups; The <i>Oncotype DX</i> RS ● Statistical significance in logistic regression analyses was determined by the likelihood ratio test; ● Area under the receiver operating characteristic curve (AUROC) was used to measure the accuracy of predictions from the logistic model; ● Expression relationships among genes were investigated through hierarchical cluster analysis using 1-Pearson r as the distance measure and pair-group average amalgamation <p>Results:</p> <ul style="list-style-type: none"> ● Clinical CR was associated with lower expression of the ER gene group and higher expression of the proliferation gene group from the 21 gene assay; ● Clinical CR was more likely in the high RS risk group (P = 0.008); ● A 50 units increase in the RS showed an OR = 5 (95% CI = 1.3-6.0); ● A14 units increase in the RS showed an OR for a clinical complete response = 1.7 (95% CI = 1.15, 2.60); ● The accuracy with which RS predicted the response to adjuvant chemotherapy throughout the range of RS values was moderate (AUROC = 0.73) | <p>The authors have established molecular profiles for sensitivity to docetaxel adjuvant chemotherapy.</p> <p>The RT-PCR technology provides a potential platform for a predictive test of chemosensitivity using small amounts of routinely processed specimens.</p> |

Evidence Table 4. Clinical utility and validity, Oncotype DX (continued)

| Study, year | Measure | Conclusions |
|------------------------------|---|--|
| Cobleigh, 2005 ⁴⁷ | <p>Context: This is one of the development studies of the Oncotype DX assay;</p> <p>End points: the likelihood of distant recurrence free survival in lymph nodes positive (> 10 nodes), ER positive and negative breast cancer patients</p> <p>Methods: Univariate and multivariate Cox proportional hazards models with stepwise variable selection and bootstrap resampling</p> <p>Results: Recurrence score classification: Low -risk group (RS < 18): 11 /78 patients (14%), rate of distant recurrence at 10 years of 29% (95% confidence interval, 0-53%) Intermediate-risk group (RS >18 and RS < 31): 19/78 patients (24%), rate of distant recurrence at 10 years of 72% (95% confidence interval, 38-88%) High-risk group (RS > 31): 48/78 patients (62%), rate of distant recurrence at 10 years of 80% (95% confidence interval, 63-89%).</p> | <p>This study found that tumor gene expression was correlated with the likelihood of distant recurrence in patients who have invasive breast cancer and more than 10 positive nodes.</p> <p>Tumor gene expression quantified by RT-PCR for 22 /203 genes was significantly correlated with distant relapse free survival (DRFS): unadjusted P-value < 0.05.</p> <p>The data generated in this study contributed to the development of the 21-gene assay and Recurrence Score algorithm.</p> |
| Esteva, 2005 ⁴⁸ | <p>Context: This study used the Oncotype DX assay and assessed its prognostic utility.</p> <p>End points (goals):</p> <ul style="list-style-type: none"> ● Assess prognosis inpatients with node-negative breast cancer who did not receive systemic therapy using formalin-fixed, paraffin-embedded tissue ● Determine whether the quantitative RT-PCR data correlated with immunohistochemistry assay data regarding estrogen receptor, progesterone receptor, and HER 2 status. <p>Methods: The Fleming and Harrington log-rank test and the Cox proportional hazards model, with the covariates age and tumor size, were used to test differences in distant recurrence among the distinct RS groups</p> <p>Results:</p> <ul style="list-style-type: none"> ● No significant correlation was found between age, tumor size, or RS and distant recurrence-free survival ● No significant difference was found between the low-, intermediate-, and high risk groups in terms of 10-year distant recurrence free survival | <p>The lack of validation of the assay observed in this study could be due to the following differences between the considered population and the NSABP B-14 population used in Paik et al 2004:</p> <ul style="list-style-type: none"> ER-positive and ER-negative patients were used; ER-positive patients were not treated with tamoxifen; An association between high nuclear grade and improved outcome was observed; Patients were from a single institution and the population could be affected by selection-bias. |

Evidence Table 4. Clinical utility and validity, Oncotype DX (continued)

| Study, year | Measure | Conclusions |
|----------------------------|---|--|
| Fan, 2006 ⁹ | <p>Context: The study is a comparison of classification agreement among 5 different gene expression based predictors (70-genes, Oncotype DX, Wound-response, Intrinsic subtypes and two-gene ratio) using the 295 samples from the consecutive cohort originally used by van de Vijver, 2002,²⁵</p> <p>Methods:</p> <ul style="list-style-type: none"> ● Cramer V statistics ● Kaplan-Meier survival analysis ● Multivariate Cox proportional hazards analysis, adjusting for age, tumor size, tumor grade, ER status, number of lymph node involved ● All analyses were performed using all the patients, as well as the ER positive patients subset (N=225) ● NB: coefficients of clinical predictors were allowed to vary between models <p>Results:</p> <ul style="list-style-type: none"> ● In multivariate Cox proportional hazards analysis all tests except the two-gene ratio were highly significant predictors of OS and DFS: 70 genes DFS HR = 3.4 (95%CI = 2.0-6.0), P value < 0.001 Two-gene ratio DFS HR = 0.91 (95%CI = 0.6-1.3), P value = 0.62 Oncotype OS HR = 4.3 (95%CI = 2.1-8.9), P value < 0.001 70 genes OS HR = 4.71 (95%CI = 2.02-11.0), P value < 0.001 Two-gene ratio OS HR = 1.00 (95%CI = 0.61-1.63), P value = 0.99 Oncotype DFS HR = 6.14 (95%CI = 1.84-20.4), P value = 0.003 ● ER status, tumor grade, tumor size, and lymph nodes were also significant predictors. ● Patients classification into dichotomized risks groups proved to be roughly similar, ranging from ~80% to ~40%: 70-genes and RS yielded a V= 0.60 70-genes and RS yielded an agreement of ~81% (239/295) | <p>Good but not perfect correlation between predictions, albeit surprising with different gene sets. Degree of prediction over and above “standard” clinical stratifiers is not clear – reclassification not done.</p> |
| Gianni, 2005 ⁴⁹ | <p>Context: This study used the Oncotype DX assay and assessed its predictive utility</p> <p>End points (goal): to directly examine the correlation between RS and pCR, and to identify additional genes associated with pCR</p> <p>Methods:</p> <ul style="list-style-type: none"> ● Univariate analysis by logistic and probit regression was used to evaluate the association of gene expression and pCR ● Forward stepwise regression and bootstrap resampling were performed to assess robustness <p>Results: The global likelihood ratio test assessing probit regression based models with and without the incorporation of the RS resulted in a P value of 0.005, showing a good discrimination of pCR by RS;</p> | <p>The authors identified a set of genes, which expression correlated with pCR to neoadjuvant doxorubicin and paclitaxel.</p> <p>The authors showed that the RS was strongly correlated with pCR.</p> |

Evidence Table 4. Clinical utility and validity, Oncotype DX (continued)

| Study, year | Measure | Conclusions |
|---------------------------|---|--|
| Habel, 2006 ⁵⁰ | <p>Context: This study used the Oncotype DX assay to assess its prognostic utility.</p> <p>End points: The risk of breast cancer-specific mortality among women with ER-positive, node-negative breast cancer. Patients were matched by age, race, year of diagnosis, and tamoxifen treatment.</p> <p>Methods:</p> <ul style="list-style-type: none"> • Conditional logistic regression analysis was performed, adjusting for tumor size and grade • Relative risks for RS, used as a continuous variable, was calculated for increments of 50 units • Standard low (RS<18), intermediate (RS=19-31), and high (RS>31) risk groups were used • Alternative RS risk stratification was obtained by RS quartiles • RS was also used as a continuous variable in the models, assessing the association with the 10-years risk of death using increments of 50 RS units <p>Results:</p> <ul style="list-style-type: none"> • Multivariate analysis with continuous RS, tumor size and grade Tamoxifen treated (55 cases and 150 controls): RR = 5.3 (95% CI = 1.6–17.2), P value = 0.003 Tamoxifen untreated (110 cases and 251 controls): RR = 2.4 (95% CI = 1.1–5.2), P value = 0.025 • Overall the patients reclassification analysis according to RS, in combination with tumor size and grade, showed that RS is able to further stratify patients' risk beyond size and grade (see below) <ul style="list-style-type: none"> 10 yr risk of death with 95% CI in the Low risk group (RS <18): <ul style="list-style-type: none"> ER+, tamoxifen treated 2.8%, 95%CI (1.7-3.9) ER+, tamoxifen untreated 6.2%, 95%CI (4.5-7.9) 10 yr risk of death with 95% CI in the Intermediate risk (RS 18-30): <ul style="list-style-type: none"> ER+, tamoxifen treated 10.7%, 95%CI (6.3-14.9) ER+, tamoxifen untreated 17.8%, 95%CI (11.8-23.3) ER+, tamoxifen treated: RR (Intermediate vs low) = 4.0 (1.8-8.8) ER+, tamoxifen untreated: RR (Intermediate vs low) = 2.7 (1.5-5.0) 10 yr risk of death with 95% CI in the High risk (RS ≥31): <ul style="list-style-type: none"> ER+, tamoxifen treated 15.5%, 95%CI (7.6-22.8) ER+, tamoxifen untreated 19.9%, 95%CI (14.2-25.2) ER+, tamoxifen treated: RR (high vs low) = 6.2 (2.4-15.8) ER+, tamoxifen untreated: RR (high vs low) = 3.3 (1.8-5.9) | <p>The RS was strongly associated with risk of breast cancer death among :</p> <ul style="list-style-type: none"> ER-positive patients treated with tamoxifen ER-positive patients not treated with tamoxifen ER-negative patients <p>Such associations remained after accounting for tumor size and grade.</p> <p>The RS was able to identify a larger subset of patients with low risk of breast cancer death than it was possible with either of these standard prognostic indicators.</p> |

Evidence Table 4. Clinical utility and validity, Oncotype DX (continued)

| Study, year | Measure | Conclusions |
|------------------------------|--|--|
| Hornberger, 2005 {#12933} | See evidence table 5 | |
| Lyman, 2007 ⁷⁵ | See evidence table 5 | |
| Mina, 2006 ⁵¹ | <p>Context: This study used the <i>Oncotype DX</i> assay to assess its predictive utility;</p> <p>End points: complete pathological response (pCR) to primary chemotherapy with anthracycline- and taxanes;</p> <p>Methods:</p> <ul style="list-style-type: none"> ● Generalized linear models with a logistic link function were used to assess the correlation of gene expression measurements with pCR. ● Logistic regression was used to examine the correlation RS with pCR. ● Logistic regression analysis was also performed to identify genes that correlated with inflammatory breast cancer (IBC). ● Correlation analyses of gene expression used Pearson linear correlation. ● Cluster analysis used 1-Pearson R squared as the distance metric and single linkage hierarchical clustering. <p>Results:</p> <ul style="list-style-type: none"> ● No correlation between <i>Oncotype DX</i> RS and pCR was demonstrated ($p = 0.67$); ● The expression of 22 individual genes, out of 274, proved to be correlate with pCR (P value < 0.05), which could be grouped in 3 functionally distinct clusters: <ul style="list-style-type: none"> Angiogenesis-related genes Proliferation-related genes Invasion related genes | <p>Although 22 genes were identified as being associated with pCR, given the large number of candidate genes tested, the likelihood of false discovery is high (13 genes were expected by chance alone).</p> <p>The 22 predictive genes contain three major predefined clusters, angiogenesis-related genes, proliferation-related genes and invasion-related genes.</p> <p>Though the <i>Oncotype DX</i> RS correlated with pCR in the INT Milan cohort of the Gianni et al study, this association was not found in the present study.</p> <p>The authors failed to find a correlation between ER and pCR.</p> |

Evidence Table 4. Clinical utility and validity, Oncotype DX (continued)

| Study, year | Measure | Conclusions |
|-------------------------------|---|---|
| Oratz, in press ⁵⁶ | <p>Context: Investigation of the impact of the Oncotype DX RS assay results on clinicians' treatment recommendations and on the actual treatment administered in patients with ER+, LN-, early (stage I or II) breast cancer;</p> <p>End points: treatment recommendations prior to the RS knowledge were compared with recommendations after the results of the assay were known and to the treatment eventually administered;</p> <p>Methods:</p> <ul style="list-style-type: none"> ● Patients were stratified by: <ul style="list-style-type: none"> Standard RS risk categories; Treatment recommendation before knowledge of the RS; Treatment recommendation post knowledge of the RS; Actual administered treatment; ● The odds ratio were computed to: <ul style="list-style-type: none"> Compare the likelihood of a change in the treatment recommendation given a low versus a high RS risk group; Evaluate the influence of the RS on the treatment actually administered. ● Exact inference methods were used for significance testing of the odds ratio; <p>Results:</p> <ul style="list-style-type: none"> ● The knowledge of the RS changed the clinicians' treatment recommendations in for 14 patients (20%): <ul style="list-style-type: none"> The odds of the physician treatment recommendation changing from hormonal therapy (HT) to chemotherapy (CT) were significantly higher ($P = 0.0011$) if the RS was high (odds 4:0) versus low (odds 1:18); The odds of the physician treatment recommendation changing from CT to HT were significantly higher ($P = 0.0340$) if the RS was low (odds 6:7) versus high (odds 0:10); ● The knowledge of the RS changed the actual administered treatment in 17 patients (25%): <ul style="list-style-type: none"> The odds of CT treatment, given an initial physician treatment recommendation for HT, were significantly higher ($P = 0.0474$) if the RS was high (odds 2:2) versus low (odds 0:19); The odds of HT treatment, given an initial physician treatment recommendation for CT, were 20.3 times higher ($P = 0.0130$) if the RS was low (odds 9:4) versus high (odds 1:9); | <p>The knowledge of the RS changed treatment recommendations and administration in patients with ER+, LN-, early breast cancer.</p> <p>An important finding is that the direction of the change was consistent with information provided by the RS.</p> |

Evidence Table 4. Clinical utility and validity, Oncotype DX (continued)

| Study, year | Measure | Conclusions |
|---------------------------|--|--|
| Paik, 2004 ^{*28} | <p>Context: This study used the Oncotype DX assay and assessed its prognostic utility.</p> <p>End points: the likelihood of distant recurrence and the Overall Survival in lymph nodes negative, ER positive breast cancer patients was considered</p> <p>Methods:</p> <ul style="list-style-type: none"> • Comparison of the high and low risk groups by Kaplan–Meier analysis with Greenwood variance estimates • Cox proportional-hazards models accounting for: <ul style="list-style-type: none"> Age (less than 50 years), clinical tumor size (less than 2cm), and RS (as continuous variable, hazard ratio for bins of 50 units); PR (as ordinal variable), ER (as ordinal variable), HER (as binary variable), age, clinical tumor size, tumor grade (1 pathologist), and RS; <p>Results:</p> <ul style="list-style-type: none"> • KM Estimates, 95% CI, and % of patients with distant recurrence at 10 years: <ul style="list-style-type: none"> Low=6.8-95%CI=4.0–9.6, 51%; Intermediate=14.3-95%CI=8.3–20.3, 22%; High =30.5-95%CI=23.6–37.4) , 27%; P<0.001 for the comparison high versus low-risk category; • Overall survival, number of patients with death event over the total number of patients at risk: <ul style="list-style-type: none"> Low=70/338; Intermediate=48/149; High =71/181; • Age, size and RS: <ul style="list-style-type: none"> p-Value=0.001, Hazard Ratio =3.21 (95%CI=(2.23–4.61) ; • PR, ER, HER, age, size, grade, and RS: <ul style="list-style-type: none"> p-Value=0.001, Hazard Ratio =2.81 (95%CI=(1.70–4.64); Similar results with more than two categories of age and tumor size (data not shown); | <p>The assay prognostic capability to quantify the likelihood of distant recurrence has been validated in tamoxifen-treated, lymph node-negative, ER-positive breast cancer.</p> |

Evidence Table 4. Clinical utility and validity, Oncotype DX (continued)

| Study, year | Measure | Conclusions |
|---------------------------|--|---|
| Paik, 2006 ^{*53} | <p>Context: This study used the Oncotype DX assay and validated the assay by assessing its predictive utility (prediction of chemotherapy benefit magnitude by RS).</p> <p>End points: freedom from distant recurrence in women with ER-positive, node-negative breast cancer from NSABP B20.</p> <p>Methods:</p> <ul style="list-style-type: none"> ● Kaplan-Meier analysis with disease recurrence free survival (DRFS). ● Multivariate Cox proportional hazards models to evaluate the interaction between chemotherapy treatment and RS as a continuous variable. ● Models included RS, individual and grouped gene expression variables, patients' age, tumor size and grade, ER status, PR status. ● RS was used as a continuous variable, age size, grade, ER and PR were dichotomized ● Models with or without RS compared by the likelihood ratio test for interaction. <p>Results:</p> <ul style="list-style-type: none"> ● Kaplan-Meier Estimates of 10 Years DRFS for Tamoxifen-Treated Patients and Tamoxifen Plus Chemotherapy-Treated Patients (10-Year DRFS (%), 95% CI, number of patients): <p>All patients :</p> <p style="padding-left: 20px;">Tamoxifen alone: 87.8, 83.3% to 92.3%, 227</p> <p style="padding-left: 20px;">Tamoxifen Plus Chemotherapy : 92.2, 89.4% to 94.9% 424</p> <p>Low risk (RS<18):</p> <p style="padding-left: 20px;">Tamoxifen alone: 96.8 93.7% to 99.9%, 135</p> <p style="padding-left: 20px;">Tamoxifen Plus Chemotherapy : 95.6 92.7% to 98.6%,218</p> <p>Intermediate risk (RS 18-30):</p> <p style="padding-left: 20px;">Tamoxifen alone: 90.9 82.5% to 99.4% 45</p> <p style="padding-left: 20px;">Tamoxifen Plus Chemotherapy : 89.1 82.4% to 95.9% 89</p> <p>High risk (RS>31):</p> <p style="padding-left: 20px;">Tamoxifen alone: 60.5 46.2% to 74.8% 47</p> <p style="padding-left: 20px;">Tamoxifen Plus Chemotherapy : 88.1 82.0% to 94.2% 117</p> | <p>The RS assay predicts the magnitude of chemotherapy benefit in women with node-negative, ER-positive breast cancer.</p> <p>If RS risk groups are considered:</p> <ul style="list-style-type: none"> a minimal benefit from chemotherapy is seen in the low risk group, however with large intervals; benefit is not assessable in the Intermediate risk group due to the uncertainty in the estimates; a large chemotherapy benefit is seen in the high risk group. |

Evidence Table 4. Clinical utility and validity, Oncotype DX (continued)

| Study, year | Measure | Conclusions |
|--------------------|---|-------------|
| Paik, 2006 (con't) | <ul style="list-style-type: none"> ● Likelihood Ratio Tests for the Interaction of Chemotherapy Treatment With the Clinical Variables and the Gene Expression Variables, Assessable NSABP B20 Patients (n = 651) RS HR = 0.32 0, 95%CI = 11 0.94 , P-value = 0.038 ● Benefit from chemotherapy usage by RS risk groups: Low risk (RS<18): RR = 1.31; 95% CI, 0.46 to 3.78, with a mean absolute decrease in distant recurrence rate at 10 years of 1.1% (SE, 2.2%) Intermediate risk (RS 18-31): RR = 0.61; 95% CI, 0.24 to 1.59, with a mean absolute increase in absolute risk of 1.8% High risk (RS>31): RR = 0.26; 95% CI, 0.13 to 0.53, with a mean absolute decrease in distant recurrence rate at 10 years of 27.6% (SE, 8.0%) ● Similar trends were observed for freedom from locoregional and/or distant recurrence and overall survival (Figs A1 and A2, online-only appendix) ● Similar results were observed when chemotherapy benefit was analyzed in the CMF and MF groups separately (data not shown); | |

RECIST = Response Evaluation Criteria in Solid Tumors; IHC = immunohistochemistry; RT-PCR = real time polymerase chain reaction; ER = estrogen receptor; PR = progesterone receptor; HER-2 = human epidermal growth factor receptor 2; AUROC=area under the receiver operating characteristic curve; CR = complete response; RS = recurrence score; OR = odds ratio; CI = confidence interval; DRFS = distant recurrence free survival; NSABP = The National Surgical Adjuvant Breast and Bowel Project; OS= overall survival; HR= hazard ratio; pCR = pathologic complete response; RR= relative risk; IBC= inflammatory breast cancer; INT= Italian National Cancer Institute of Milan, Italy; LN= lymph node; HT = hormonal therapy; CT = chemotherapy; KM= Kaplan Meier; SE = standard error; CMF= cyclophosphamide, methotrexate, and fluorouracil; MF= methotrexate and fluorouracil.

Evidence Table 5. Critical appraisal of economic analyses of breast cancer gene expression testing

| Dimension of Quality | Questions for Critical Appraisal ⁴¹ | Answers | | |
|--|--|---|--|--|
| | | Hornberger, 2005 ¹⁰⁰ | Lyman, 2007 ¹⁰¹ | Oestreicher, 2005 ⁷⁶ |
| Structure | | | | |
| Statement of decision problem or objective | Is there a clear statement of the decision problem? | Yes; Is a 21-Gene RT-PCR assay economically and clinically useful in assisting in classification of ER+, LN- early stage breast cancer patients and in determination of chemotherapy? | Yes; Is the use of a 21-Gene RT-PCR assay to assist in the risk classification and chemotherapy recommendation of early-stage breast cancer patients clinically and economically beneficial? | Yes; Is using the Netherlands Cancer Institute GEP assay versus NIH guidelines cost-effective for the identification of early stage breast cancer patients who would benefit from adjuvant chemotherapy? |
| | Is the objective of the evaluation and of the model specified and consistent with the stated decision problem? | Yes; to estimate the incremental benefits, costs, and cost-effectiveness of using a recurrence score based on the 21-Gene RT-PCR assay that would assign risk more accurately and help guide the decision to prescribe chemotherapy | Yes; to incorporate and extend validation results from previous studies of NSABP B-14 and B-20 into an economic model of the 21-gene expression signature to guide the use of adjuvant chemotherapy and hormonal therapy in patients with ER+, LN- early stage breast cancer | Yes; to assess the incremental cost and QALYs associated with use of the GEP assay compared to use of the NIH guidelines. |
| | Is the primary decision-maker specified? | Yes; physician | Yes; physician | Yes; physician |

Evidence Table 5. Critical appraisal of economic analyses of breast cancer gene expression testing (continued)

| Dimension of Quality | Questions for Critical Appraisal | Answers | | |
|-----------------------------------|---|--|---|--|
| | | Hornberger, 2005 | Lyman, 2007 | Oestreicher, 2005 |
| Structure (continued) | | | | |
| Statement of scope or perspective | Is the perspective for the model stated clearly? | Yes; Authors state that the model takes a “societal perspective considering survival, quality of life, and relevant costs.” | No. | Yes; Authors state that the model takes a “societal perspective for this analysis and follows the recommendations of the Panel on Cost-Effectiveness in Health and Medicine.” |
| | Are the model inputs consistent with the stated perspective? | Not entirely; the model does not include all costs that are relevant to a societal perspective, and specifically does not include direct non-medical costs (other than patient time) and indirect costs. | Not applicable since perspective is not stated. | Yes; the model includes costs that are relevant to a societal perspective, including direct medical costs and non-medical costs (e.g., lost wages from time spent traveling to treatment facilities and in treatment). |
| | Has the scope of the model been stated and justified? | Partly; Authors state the model includes survival, quality of life, and relevant costs, and forecasts outcomes over the patient’s lifetime. Authors do not justify exclusion of some types of societal costs, and do not justify using a timeframe beyond the 10-year period for which recurrence data is available. | Partly; Authors indicate that the model includes clinical, economic, and quality of life outcomes, and projects the impact on life expectancy. Authors do not justify exclusion of some types of costs, and do not justify using a timeframe beyond the 10-year period for which recurrence data was available. | Yes; Authors state the model includes survival, quality of life, and relevant costs, and forecasts outcomes over the patient’s lifetime. However, the model is based on a 6.7-year period for which recurrence data is available yet is projected over the patients’ lifetime. |
| | Are the outcomes of the model consistent with the perspective, scope, and overall objective of the model? | Yes; Outcomes include survival in life-years, QALYs, and overall costs for breast cancer. | Yes; Outcomes include life-years saved, QALYs, and costs of cancer care. | Yes; Outcomes include expected proportion of distant recurrences prevented, QALYs, and overall costs for breast cancer. |

Evidence Table 5. Critical appraisal of economic analyses of breast cancer gene expression testing (continued)

| Dimension of Quality | Questions for Critical Appraisal | Answers | | |
|------------------------------|---|--|--|---|
| | | Hornberger, 2005 | Lyman, 2007 | Oestreicher, 2005 |
| Structure (continued) | | | | |
| Rationale for structure | Is the structure of the model consistent with a coherent theory of the health condition under evaluation? | Yes; The authors use a Markov model with transition states for: no recurrence; recurrence without metastatic progression; recurrence with metastatic progression; and death. | Uncertain; The authors do not provide sufficient information about the structure of the model. The model seems to be based on comparing outcomes for 3 adjuvant treatment strategies. Patients are classified based on RS score and then a treatment strategy is applied: Strategy 1 (all patients receive tamoxifen), Strategy 2 (all patients receive chemotherapy and tamoxifen), Strategy 3 (if RS classification = low risk, then only tamoxifen; if RS classification = intermediate or high risk, then chemotherapy and tamoxifen). | Yes; The authors use a Markov model with transition states for: no evidence of disease; distant recurrence; and death. |
| | Are the sources of data used to develop the structure of the model specified? | Yes | Uncertain because the structure of the model is not specified | Yes |
| | Are the causal relationships described by the model structure justified appropriately? | Yes; Authors use clinical data to determine recurrence rates and assume that all high risk patients will receive chemotherapy and all low risk patients will not receive chemotherapy. | Uncertain because the structure of the model is not specified | Yes; Authors use clinical data to determine recurrence rates and assume that all women classified as "poor prognosis" will receive chemotherapy and all women classified as "good prognosis" will not receive chemotherapy. |
| Structural Assumptions | Are the structural assumptions transparent and justified? | Yes | No | Yes |
| | Are the structural assumptions reasonable given the overall objective, perspective, and scope of the model? | Yes | Uncertain | Yes |

Evidence Table 5. Critical appraisal of economic analyses of breast cancer gene expression testing (continued)

| Dimension of Quality | Questions for Critical Appraisal | Answers | | |
|------------------------------|--|---|--|--|
| | | Hornberger, 2005 | Lyman, 2007 | Oestreicher, 2005 |
| Structure (continued) | | | | |
| Strategies/comparators | Is there a clear definition of the options under evaluation? | Yes; NCCN system as current standard versus the RT-PCR assay | Yes; clear description of treatment strategies as noted above | Yes; NIH clinical guidelines as the standard versus the Netherlands Cancer Institute GEP assay. |
| | Have all feasible and practical options been evaluated? | No; This analysis focuses only on comparing use of one test, the RT-PCR assay, to use of the NCCN guidelines. | No: This analysis focuses only on comparing outcomes of treatment guided by one test, the 21-gene expression signature, to outcomes associated with using tamoxifen alone or chemotherapy followed by tamoxifen. | No; This analysis focuses only on comparing use of one test, the Netherlands Cancer Institute GEP assay, to use of the NIH guidelines. A stated limitation by the authors is that other guidelines (e.g., NCCN) are available but were not used in this comparison. |
| | Is there justification for the exclusion of feasible options? | No | No | Partly; the authors state that the NCCN criteria are more conservative; they would have yielded higher cost and would have had higher sensitivity than the NIH guidelines. They suggest that using NCCN criteria would be unlikely to change the outcome of the model. |
| Model type | Is the chosen model type appropriate given the decision problem and specified causal relationships within the model? | Yes | Uncertain | Yes |

Evidence Table 5. Critical appraisal of economic analyses of breast cancer gene expression testing (continued)

| Dimension of Quality | Questions for Critical Appraisal | Answers | | |
|------------------------------|--|--|---|--|
| | | Hornberger, 2005 | Lyman, 2007 | Oestreicher, 2005 |
| Structure (continued) | | | | |
| Time horizon | Is the time horizon of the model sufficient to reflect all important differences between options? | Yes | Yes | Yes |
| | Are the time horizon of the model, the duration of the treatment, and the duration of the treatment effect described and justified? | Yes; The model forecasts outcomes over the patient's lifetime and assumes treatment time of 6 months, but does not explicitly state any assumption regarding the duration of the treatment effect. | Yes; The model estimates the impact on life expectancy and presents the assumed duration of treatment options, the longest of which was 24 weeks. | Yes; The model forecasts outcomes over the patient's lifetime and assumes initial treatment time of 6 months, but does not explicitly state any assumption regarding the duration of the treatment effect. |
| Disease states/ pathways | Do the disease states (state transition model) or the pathways (decision tree model) reflect the underlying biological process of the disease in question and the impact of interventions? | Yes | Uncertain | Yes |
| Cycle length | Is the cycle length defined and justified in terms of the natural history of disease? | Yes; Cycle length is defined as 1 year. | No; Cycle length is not specified. | Yes; Cycle length is defined as 1 year. |

Evidence Table 5. Critical appraisal of economic analyses of breast cancer gene expression testing (continued)

| Dimension of Quality | Questions for Critical Appraisal | Answers | | |
|----------------------|--|---|--|---|
| | | Hornberger, 2005 | Lyman, 2007 | Oestreicher, 2005 |
| Data | | | | |
| Data identification | Are the data identification methods transparent and appropriate given the objectives of the model? | Yes; The studies and cases used in the model are clearly referenced. | Yes; The authors cite previous work for some of the details. | Yes; The studies and assumptions used in the model are clearly referenced. |
| | Where choices have been made between data sources, are these justified appropriately? | Yes; Only NSABP B-14 data are used. This is justified by the limited amount of comparable sources of data with adequate followup. | Yes; The authors explain their use of data from NSABP B-20 and NSABP B-14. | Yes; The Netherlands Cancer Institute cohort data are used. |
| | Has particular attention been paid to identifying data for the important parameters in the model? | Yes; Authors focus on: 1) propensity to change chemotherapy decision based on recurrence score, 2) relative risk reduction of chemotherapy and tamoxifen versus tamoxifen only, 3) cost of chemotherapy, and 4) proportion of patients defined by NCCN as low risk. | Uncertain given the lack of information about the structure of the model | Yes; Authors focus on propensity to make chemotherapy decision based on the classification of low risk versus high-risk, relative risk reduction of chemotherapy, cost of chemotherapy, and proportion of patients defined by NIH criteria as low risk. |
| | Has the quality of the data been assessed appropriately? | Not reported | Not reported | Not reported |
| | Where expert opinion has been used, are the methods described and justified? | Not applicable | Not applicable | Not applicable |
| Data modeling | Is the data modeling methodology based on justifiable statistical and epidemiological techniques? | Yes | Uncertain given the lack of information about the structure of the model | Yes |

Evidence Table 5. Critical appraisal of economic analyses of breast cancer gene expression testing (continued)

| Dimension of Quality | Questions for Critical Appraisal | Answers | | |
|-------------------------|--|--|--|--|
| | | Hornberger, 2005 | Lyman, 2007 | Oestreicher, 2005 |
| Data (continued) | | | | |
| Baseline Data | Is the choice of baseline data described and justified? | Yes; Percent chance of recurrence and treatment plans are cited and described. | Yes; Table 1 describes most of the baseline data. Other values are described throughout the paper | Yes; Probabilities of recurrence and treatment plans are described and justified. |
| | Are transition probabilities calculated appropriately? | Yes; Probabilities appear to be obtained from prior meta-analyses. | Uncertain, given the lack of information about the structure of the model. However, probabilities are taken from previous analysis of the 21-Gene RT-PCR assay | Yes; Probabilities were estimated from empirical analyses performed on patient level data from the Netherlands Cancer Institute cohort, which was also the population for a validation study of GEP. |
| | Has a half-cycle correction been applied to both cost and outcome? | Not reported | Not reported | Not reported |
| | If not, has this omission been justified? | No | No | No |

Evidence Table 5. Critical appraisal of economic analyses of breast cancer gene expression testing (continued)

| Dimension of Quality | Questions for Critical Appraisal | Answers | | |
|-------------------------|---|---|--|--|
| | | Hornberger, 2005 | Lyman, 2007 | Oestreicher, 2005 |
| Data (continued) | | | | |
| Treatment effects | If relative treatment effects have been derived from trial data, have they been synthesized using appropriate techniques? | Yes; The authors use meta-analyses of clinical trials to estimate the relative risk reduction associated with chemotherapy plus tamoxifen versus tamoxifen alone. | Uncertain; Table 1 suggests that a single study is used as the source of estimates for relative effects of tamoxifen, chemotherapy plus tamoxifen, and RS-guided therapy. | Yes; The authors use the meta-analysis from the Early Breast Cancer Trialists' Group trial to estimate the relative risk reduction associated with chemotherapy. |
| | Have the methods and assumptions used to extrapolate short-term results to final outcomes been documented and justified? | Partly; The authors assume a constant recurrence rate over 10 years, but do not justify extrapolation beyond 10 years. | Not reported | Partly; The authors assume a constant recurrence rate over 6.7 years, but do not justify extrapolation beyond 6.7 years. |
| | Have alternative assumptions been explored through sensitivity analysis? | Yes, using sensitivity analysis and Monte Carlo simulation | Uncertain; Ranges for costs, healthy life expectancy, and utility with chemotherapy are given in Table 1. Although the authors do not state whether a sensitivity analysis was performed for these assumptions in particular, they indicate that a Monte Carlo analysis was performed. | Yes, a wide variety of assumptions has been explored using sensitivity analysis and Monte Carlo simulation |
| | Have assumptions regarding the continuing effect of treatment once treatment is complete been documented and justified? Have alternative assumptions been explored through sensitivity analysis? | Yes | Yes; The introduction also mentions ongoing investigations into the long-term effects of chemotherapy, particularly neurocognitive dysfunction and secondary malignancies | Yes |

Evidence Table 5. Critical appraisal of economic analyses of breast cancer gene expression testing (continued)

| Dimension of Quality | Questions for Critical Appraisal | Answers | | |
|-------------------------------------|---|---|---|--|
| | | Hornberger, 2005 | Lyman, 2007 | Oestreicher, 2005 |
| Data (continued) | | | | |
| Quality of life weights (utilities) | Are the utilities incorporated into the model appropriate? | Yes; Table 3 lists several health-state utilities that are appropriate to include in the analysis. One limitation is that the sensitivity analysis considered only a relatively narrow range of estimates for each utility value. | Uncertain: The only health-state utility mentioned in this report is for "utility with chemotherapy." No mention is made of any other relevant health-state utilities. | Yes; Table 1 listed several utilities that are appropriate to include in the analysis. These included: utilities for no adjuvant chemotherapy; adjuvant chemotherapy; no evidence of disease and distant recurrence |
| | Are the methods of derivation for the utilities justified? | Uncertain; The authors cite several references from peer-reviewed journals as the sources for the utility estimates, but they do not discuss any of the potential limitations of how those utilities were originally derived. | Partly: The authors report that they used a time trade-off method to estimate the impact of treatment strategies on health-related quality of life, but they provide no details about how the estimates were derived. | Unclear; The authors discuss their derivation of utilities and cite several references from peer-reviewed journals as the sources for the utility estimates. But they gave no explanation of the methods or scaling techniques that were used to derive the utility estimates. The utility values generally seemed very low when compared to the values typically assigned to serious health conditions. |
| Costs | Are the costs incorporated into the model justified? | Yes, but see the comment above regarding exclusion of some types of societal costs | Yes, but they provide no details about how the estimates were derived. | Yes, although it is unclear whether all relevant societal costs have been included. |
| | Has the source for all costs been described? | Yes | Yes; Most costs have been cited in Table 1. A reference is given for additional details. | Yes |
| | Have discount rates been described and justified given the target decision-maker? | Yes; The investigators discount both costs and survival by an annual rate of 3%. | Not reported | Yes; The investigators discounted costs and QALYs by an annual rate of 3%. |

Evidence Table 5. Critical appraisal of economic analyses of breast cancer gene expression testing (continued)

| Dimension of Quality | Questions for Critical Appraisal | Answers | | |
|-------------------------|---|--|---|--|
| | | Hornberger, 2005 | Lyman, 2007 | Oestreicher, 2005 |
| Data (continued) | | | | |
| Data incorporation | Have all data incorporated into the model been described and referenced in sufficient detail? | Yes; Data inputs have been referenced and most have been described in sufficient detail. | No; The individual costs and recurrence probabilities have been justified and referenced. However, not all data inputs have been described. | Yes; Data inputs have been referenced and have been described in sufficient detail. |
| | Has the use of mutually inconsistent data been justified (i.e., are assumptions and choices appropriate?) | Not applicable | Not applicable | Not applicable |
| | Is the process of data incorporation transparent? | Yes; The incorporation of most costs into the model are described. However, calculation of the "follow-up" number is not described, although it seems to be incorporating treatment side-effects, monitoring for side-effects and end-of-life costs. | No; Not enough information is given to understand how the authors derived the estimates of "Costs" and "LYS" in Table 2 and "Expected Costs" from Figure 2. | Yes; The incorporation of costs into the model is described. |
| | If data have been incorporated as a distribution, has the choice of distribution for each parameter been described and justified? | Generally, point estimates are used, not distributions. As part of the sensitivity analysis, the investigators used a Monte Carlo probabilistic analysis, and referred to a technical appendix for the details. | Yes; The authors use a figure to show the distribution of distant recurrence estimates by RS. | Yes. The investigators conducted one-way sensitivity analysis varying the values of probabilities, costs and utilities on model results. As part of the sensitivity analysis, the investigators used a Monte Carlo probabilistic analysis. |
| | If data have been incorporated as distributions, is it clear that second order uncertainty is reflected? | Not reported | Not reported | Yes. |

Evidence Table 5. Critical appraisal of economic analyses of breast cancer gene expression testing (continued)

| Dimension of Quality | Questions for Critical Appraisal | Answers | | |
|---------------------------|--|--|---|---|
| | | Hornberger, 2005 | Lyman, 2007 | Oestreicher, 2005 |
| Data (continued) | | | | |
| Assessment of uncertainty | Have methodological uncertainties been addressed by running alternative versions of the model with different methodological assumptions? | Not reported | Not reported | Yes. A number of alternative testing hypotheses and alternative testing strategies were considered in the model. |
| | Is there evidence that structural uncertainties have been addressed via sensitivity analysis? | Not reported | Not reported | Not reported. |
| | Has heterogeneity been dealt with by running the model separately for different subgroups? | Yes; The model considers NCCN low risk and NCCN high risk groups, and is applied to a cohort of 100 patients similar to those in the B-14 trial. | Yes; The model considers the B-14 and B-20 studies/ | Yes; The model considers low risk and high risk groups, and is applied to a hypothetical cohort of patients similar to those in the Netherlands Cancer Institute cohort. Additional prognostic groups using different combinations of GEP and NIH results were also considered. |
| | Are the methods of assessment of parameter uncertainty appropriate? | Yes | Yes | Yes |
| | If data are incorporated as point estimates, are the ranges used for sensitivity analysis stated clearly and justified? | Yes; The ranges are stated in Table 3, but are not fully justified. | Yes; The ranges for most data points are clearly stated in Table 1. | Yes; The ranges are stated in Table 1, but are not fully justified. |

Evidence Table 5. Critical appraisal of economic analyses of breast cancer gene expression testing (continued)

| Dimension of Quality | Questions for Critical Appraisal | Answers | | |
|----------------------|---|------------------|----------------|--|
| | | Hornberger, 2005 | Lyman, 2007 | Oestreicher, 2005 |
| Consistency | | | | |
| Internal consistency | Is there evidence that the mathematical logic of the model has been tested thoroughly before use? | Not reported | Not reported | Yes; the analysis was performed using a Microsoft Excel spreadsheet and then validated using decision analysis software. |
| External consistency | Are any counterintuitive results from the model explained and justified? | Not applicable | Not applicable | Not applicable |
| | If the model has been calibrated against independent data, have any differences been explained and justified? | Not reported | Not reported | Not reported |
| | Have the results of the model been compared with those of previous models and any differences in results explained? | Not reported | Not reported | Not reported |

RT-PCR= reverse transcriptase polymerase chain reaction; ER=estrogen receptor; LN= lymph node; NSABP= National Surgical Adjuvant Breast and Bowel Project; GEP= gene expression programming; NIH= National Institute of Health; QALY= quality-adjusted life-year; RS= recurrence score; NCCN= National Comprehensive Cancer Network; LYS= life years saved.

Evidence Table 6. Study design, MammaPrint®

| Study, Year | Country | Study (data collection) period | Study Type | Population size, N | Blinded (Y/N) | Study purpose |
|------------------------------------|--------------------|---------------------------------------|-------------------|--|----------------------|----------------------------------|
| Ach, 2007 ⁵⁷ | | | | | | Inter-laboratory reproducibility |
| Buyse, 2006 ⁵⁹ | France, Sweden, UK | 1980-1998 | retrospective | Clinical low risk/gene low risk, 52 | NR | Validation |
| | | | | Clinical low risk/gene high risk, 28 | | |
| | | | | Clinical high risk/gene low risk, 59 | | |
| | | | | Clinical high risk/gene high risk, 163 | | |
| Fan, 2006 ⁷⁹ | Netherlands | NR | retrospective | Poor prognosis, 180 | NA | Validation |
| | | | | Good prognosis, 115 | | |
| Glas, 2006 ⁵⁸ | Netherlands | NR | retrospective | 162 | NR | Development |
| van de Vijver, 2002 ^{25*} | Netherlands | 1984-1995 | retrospective | Poor prognosis, 180 | NR | Validation |
| | | | | Good prognosis, 115 | | |
| van't Veer ^{21*} | Netherlands | 1983-1996 | retrospective | 98 (they say 98, but they are 97!!!!) | NR | Development and validation |

* Supplemental information

NR = not reported; NA = not applicable

Evidence Table 7. Study population characteristics, MammaPrint®

| Study, Year | Intervention | General Characteristics | Diagnosis (es) [include details] | Treatments and Outcomes |
|---------------------------|--|--|---|---|
| Ach, 2007 ⁵⁷ | There are not patient characteristics reported in this article. This is a study of analytic validity | | | |
| Buyse, 2006 ⁵⁹ | Clinical low risk/gene low risk§ n=52 | Age, n (%) < 41, 8 (15) 41-50, 36 (69) 51-60, 8 (15) Menopausal status: NR Race: NR Exclusion criteria: <ul style="list-style-type: none"> • >61yearsOld • >T1-T2(>5cm) • previous malignancy (except basal cell carcinoma) • bilateral synchronous breast | TS <1 cm, 5 (9) TS 1-2 cm, 30 (58) TS 2-5 cm, 17 (33) TG good diff, 22 (43) TG Int diff, 29 (56) TG poor diff, 0 (0) TG UK, 1 (2) ER+, 52 (100) ER-, 0 (0) Diagnostic method: clinical and microarray | Treatment: NR Outcome: Proportion alive at 10y (95% CI), 0.88 (0.74 to 0.95) |
| | Clinical low risk/gene high risk n=28 | Age, n (%) < 41, 1 (4) 41-50, 22 (78) 51-60, 5 (18) Menopausal status: NR Race: NR Exclusion criteria: see clinical low risk/gene low risk | TS <1 cm, 4 (14) TS 1-2 cm, 12 (43) TS 2-5 cm, 12 (43) TG good diff, 12 (43) TG Int diff, 16 (57) TG poor diff, 0 (0) TG UK, 0 (0) ER+, 28 (100) ER-, 0 (0) Diagnostic method: clinical and microarray | Treatment: NR Outcome: Proportion alive at 10y (95% CI), 0.69 (0.45 to 0.84) |
| | Clinical high risk/gene low risk§ n=59 | Age, n (%) < 41, 9 (15) 41-50, 20 (34) 51-60, 30 (51) Menopausal status: NR Race: NR Exclusion criteria: see clinical low risk/gene low risk | TS <1 cm, 0 (0) TS 1-2 cm, 17 (29) TS 2-5 cm, 42 (71) TG good diff, 7 (12) TG Int diff, 41 (69) TG poor diff, 11 (18) TG UK, 0 (0) ER+, 54 (91) ER-, 5 (9) Diagnostic method: clinical and microarray | Treatment: NR Outcome: Proportion alive at 10y (95% CI), 0.89 (0.45 to 0.95) |

Evidence Table 7. Study population characteristics, MammaPrint (continued)

| Study, Year | Intervention | General Characteristics | Diagnosis (es) [include details] | Treatments and Outcomes |
|-------------------------|---|--|--|---|
| Buyse, 2006 (cont') | Clinical high risk/gene high risk§ n=163 | Age, n (%) < 41, 40 (24) 41-50, 67 (41) 51-60, 56 (34) Menopausal status: NR Race: NR Exclusion criteria: see clinical low risk/gene low risk | TS <1 cm, 2 (1) TS 1-2 cm, 40 (24) TS 2-5 cm, 121 (74) TG good diff, 6 (3) TG Int diff, 39 (23) TG poor diff, 113 (69) TG UK, 5 (3) ER+, 78 (48) ER-, 85 (52) Diagnostic method: clinical and microarray | Treatment: NR Outcome: Proportion alive at 10y (95% CI), 0.69 (0.61 to 0.76) |
| Fan, 2006 ⁹⁸ | Poor prognosis | Age, n (%) <40yrs 52 (29) 40-44yrs, 41 (23) 45-49yrs, 55 (31) ≥50yrs; 32 (18) Menopausal status: NR Race: NR Exclusion criteria: <ul style="list-style-type: none"> • Previous history of cancer (except non-melanoma skin cancer) • Apical axillary lymph node involvement • Age at diagnosis >52 years • Tumor >5 cm in diameter at pathological examination | H I (good); 19/180 (11%) H II (intermediate); 56/180 (31%) H III (poor); 105/180 (58%) TS ≤20mm; 84/180 (47%) TS >20mm; 96/180 (53%) LN# 0; 91/180 (51%) LN#1-3; 63/180 (35%) LN# ≥4; 26/180 (14%) ER -; 66/180 (37%) ER +; 114/180 (63%) VI Absent; 108/180 (60%) VI 1-3 vessels; 18/180 (10%) VI >3 vessels; 54/180 (30%) Diagnostic method: RT-PCR | Treatment: <ul style="list-style-type: none"> h Yes; 23/180 (13%) h No; 157/180 (87%) a Yes; 66/180 (37%) a No; 114/180 (63%) s Breast conserving; 97/180 (54%) s mastectomy; 83/180 (46%) Outcome: NR |

Evidence Table 7. Study population characteristics, MammaPrint (continued)

| Study, Year | Intervention | General Characteristics | Diagnosis (es) [include details] | Treatments and Outcomes |
|-------------------------------------|--|---|--|--|
| Fan, 2006 (cont') | Good prognosis | Age, n (%) <40yrs, 11 (10) 40-44yrs, 44 (38) 45-49yrs, 43 (37) ≥50yrs, 17 (15) Menopausal status: NR Race: NR Exclusion criteria: see poor prognosis | H I (good); 56/115 (49%) H II (intermediate); 45/115 (39%) H III (poor); 14/115 (12%) TS ≤20mm; 71/115 (62%) TS >20mm; 44/115 (38%) LN# 0; 60/115 (52%) LN# 1-3; 43/115 (37%) LN# ≥4; 12/115 (10%) ER-; 3/115 (3%) ER +; 112/115 (97%) VI Absent; 77/115 (67%) VI 1-3 vessels; 12/115 (10%) VI >3 vessels; 26/115 (23%) Diagnostic method: RT-PCR | Treatment: h: Yes; 17/115 (15%) h: No; 88/115 (85%) a: Yes; 44/115 (38%) a: No; 71/115 (62%) s: Breast conserving; 64/115 (56%) s: mastectomy; 51/115 (44%) Outcome: NR |
| Glas, 2006 ⁵⁸ | There are not patient characteristics reported in this article. This is a study of analytic validity | | | |
| van de Vijver, 2002 ^{25 a} | Poor prognosis | Age, n (%) <40yrs; 52 (29) 40-44yrs; 41 (23) 45-49yrs; 55 (31) ≥50yrs; 32 (18) Menopausal status: NR Race: NR Exclusion criteria: <ul style="list-style-type: none"> • Previous history of cancer (except non-melanoma skin cancer) • Apical axillary lymph node involvement • Age at diagnosis >52 years • Tumor >5 cm in diameter at pathological examination | H I (good); 19/180 (11%) H II (intermediate); 56/180 (31%) H III (poor); 105/180 (58%) TS ≤20mm; 84/180 (47%) TS >20mm; 96/180 (53%) LN# 0; 91/180 (51%) LN#1-3; 63/180 (35%) LN# ≥4; 26/180 (14%) ER -; 66/180 (37%) ER +; 114/180 (63%) VI Absent; 108/180 (60%) VI 1-3 vessels; 18/180 (10%) VI >3 vessels; 54/180 (30%) | Treatment: h: Yes; 23/180 (13%) h: No; 157/180 (87%) a: Yes; 66/180 (37%) a: No; 114/180 (63%) s: Breast conserving; 97/180 (54%) s: mastectomy; 83/180 (46%) Outcome: NR |

Evidence Table 7. Study population characteristics, MammaPrint (continued)

| Study, Year | Intervention | General Characteristics | Diagnosis (es) [include details] | Treatments and Outcomes |
|---|----------------|---|--|--|
| van de Vijver, 2002 (cont') | Good prognosis | Age, n (%) <40yrs; 11 (10) 40-44yrs; 44 (38) 45-49yrs; 43 (37) ≥50yrs; 17 (15) Menopausal status: NR Race: NR Exclusion criteria: see poor prognosis | H I (good); 56/115 (49%) H II (intermediate); 45/115 (39%) H III (poor); 14/115 (12%) TS ≤20mm; 71/115 (62%) TS >20mm; 44/115 (38%) LN# 0; 60/115 (52%) LN# 1-3; 43/115 (37%) LN# ≥4; 12/115 (10%) ER-; 3/115 (3%) ER +; 112/115 (97%) VI Absent; 77/115 (67%) VI 1-3 vessels; 12/115 (10%) VI >3 vessels; 26/115 (23%) | Treatment: h: Yes; 17/115 (15%) h: No; 88/115 (85%) a: Yes; 44/115 (38%) a: No; 71/115 (62%) s: Breast conserving; 64/115 (56%) s: mastectomy; 51/115 (44%) Outcome: NR |
| ^a Poor and Good prognosis groups using the optimized sensitivity threshold from van't Veer 2002 ²¹ | | | | |
| van't Veer, 2002 ^{21 b} | | Mean age, y (range), SD: 44.9 (28-54), 6.9 (for the 78 patients used to develop the test + 19 of the validation; BRCA1 patients non included) Menopausal status: NR Race: NR Exclusion criteria: <ul style="list-style-type: none"> • Tumor larger than 5cm (>T2) • metastases (M1) • Older than 55yrs at diagnosis • previous malignancy • not diagnosed between 1983 and 1996 | H I (good); 11/97 (11.3%) H II (intermediate); 26/97 (26.8%) H III (poor); 60/97 (61.8%) TS ≤20mm; 57/97 (58.8%) TS >20mm; 40/97 (41.2%) LN# 0; 97/97 (100%) ER -; 29/97 (29.8%) (by IHC, <10% positive nuclei) ER +; 68/97 (70.2%) (by IHC, >10% positive nuclei) PR -; 41/97 (42.3%) (by IHC, <10% positive nuclei) PR +v; 56/97 (57.7%) (by IHC, >10% positive nuclei) VI Absent; 69/97 (71.2%) VI Present 28/97 (28.8%) LI Absent; 18/97 (18.5%) LI Present 79/97 (81.5%) | Treatment: h: hormonal therapy (3 patients) a: Adjuvant Chemotherapy (3 patients) s: Modified radical mastectomy (35 patients) s: Breast conserving treatment (72 patients) S: axillary lymph-node dissection (all 97 patients) Outcome: Metastases within 5yrs from surgery: 51 patient did NOT develop metastases ; 46 patient developed metastases |
| ^b Development (78 patients): classified in the Poor group=35 ; classified in the Good group=43 ; Validation (19 patients): classified in the Poor group=5 ; classified in the Good group=14 with the optimized sensitivity threshold | | | | |

Evidence Table 7. Study population characteristics, MammaPrint (continued)

* Diagnoses: H = histology/histologic type; TS = tumor size/diameter; LN# = # of Lymph nodes at primary diagnosis; LN = lymph node status; TG = tumor grade; ER = estrogen receptor status; HER = HER neu status; PR = Progesterone receptor status; SP = S-phase fraction; Pl = ploidy; VI = vascular invasion; PTS = site of primary tumor; HR = hormonal status, ER or PR or both; DR = distant recurrence; LI lymphotic infiltrate

† Treatments: H = hormonal; A = adjuvant; S = surgical; R = radiation

‡ Treatment outcomes: OS = overall survival; RFS = relapse-free survival (first disease relapse, irrespective of site); CR = complete response; PR = partial response; RS: recurrence score; SD = stable disease; PD = progressive disease; PCR = pathologic complete response; D = death

§ Data not reported in article, referred directly to van de Vijver, 2002²⁵.

|| Data not reported in article, referred directly to van't Veer, 2002²¹.

NR= not reported; CI= confidence interval; RT-PCR= reverse transcriptase polymerase chain reaction.

Evidence Table 8. Analytic validity, MammaPrint®

| Study, year | Measure | Conclusions |
|---------------------------|--|---|
| Ach, 2007, ⁵⁷ | <p>Context: MammaPrint assay was used to evaluate the intra- and inter-laboratory reproducibility of the assay involving three laboratories. Variation in RNA amplification and labeling, hybridization and wash, and slide scanning was measured on 4 tumors, dye-flip design, 24 slides (8 per site).</p> <p>Methods: To assess reproducibility in this study, ANOVA P values and Pearson's correlation were used.</p> <p>Results:</p> <ul style="list-style-type: none"> ● Replicate hybridizations Pearson's correlation at the same site: <ul style="list-style-type: none"> For 1 tumor in 1 sub-array = 0.983 For 2 tumor in 2 sub-arrays = 0.988 For all the other technical replicates > 0.993 ● Scanning reproducibility across sites: <ul style="list-style-type: none"> Cy3: Pearson correlation >0.995, slope = 0.97 Cy5: less reproducible (data not shown) ● 70-gene signature reproducibility: <ul style="list-style-type: none"> No differences between hybridization sites No differences between hybridization days (regardless of site) Statistically significant difference (P value <0.05) between labeling sites for two tumors | <p>The authors showed very low influence on sample-to-reference ratios based on averaged triplicate measurements in the two-color experiments;</p> <p>RNA labeling was the largest contributor to inter-laboratory variation;</p> <p>Overall, despite this variation, measurement of 70-gene signature in three different laboratories was found to be highly robust;</p> |
| Buyse, ⁵⁹ 2006 | <p>Context: The MammaPrint assay was used. Out of the 403 eligible samples the analysis was carried on 326 samples.</p> <p>Methods: Standard MammaPrint assay protocols were used, using fresh-frozen specimens</p> <p>Results:</p> <ul style="list-style-type: none"> ● 326 patients out of 403 successfully analyzed, for the following reasons: <ul style="list-style-type: none"> Poor RNA yield: 77/403 patients (19.1%) Successful assays: 326/403 patients (80.9%) | |

Evidence Table 8. Analytic validity, MammaPrint (continued)

| Study, year | Measure | Conclusions |
|--------------------------|--|--|
| Glas, 2006 ⁵⁸ | <p>Context: MammaPrint assay development through re-analysis of patients from the van't Veer²¹ and van de Vijver²⁵ cohorts for which an RNA aliquot or the frozen specimen were available. A different reference RNA was used, as well as a different quantification method.</p> <p>Methods:</p> <ul style="list-style-type: none"> • 162 total samples from fresh-frozen specimens: <ul style="list-style-type: none"> 84 patients from the van de Vijver²⁵ cohort All 78 patients form the van't Veer cohort²¹ A combination of the two population above: 145 (84+61) LN-negative patients Re-analysis on 49 patients • Dye-swap hybridization over reference RNA from 105 sample from the van de Vijver series (balanced in terms of Good and Poor) • Relative gene expression quantified as the error-weighted average over triplicate spots of the log₁₀ ratio • Pearson's correlation to assess correlation with original data and reproducibility • ANOVA analysis to model variability in repeated experiments using the 70 genes of the signature • Reproducibility in time was assessed by repeated measurements of RNA aliquots: <ul style="list-style-type: none"> 1 patient with cosine correlation to Good profile = 0.61, for 12 months 1 patient with cosine correlation to Good profile = -0.44, for 12 months 1 border-line with cosine correlation to Good profile = 0.43, for 4 months <p>Results:</p> <ul style="list-style-type: none"> • Comparison to original 70-gene signature data, Pearson's Correlations and in repeated measurements: <ul style="list-style-type: none"> 78 van't Veer²¹ patients, $r = 0.92$, $p \text{ value} < 0.0001$ 145 (84+61) van de Vijver²⁵ LN-negative patients: $r = 0.88$ $p \text{ value} < .0001$ 49 patients analyzed twice, $r = 0.995$ • Reproducibility results from ANOVA analysis: <ul style="list-style-type: none"> No variation within individuals ($p \text{ value} = 0.96$) Significant variation between individuals and genes • Reproducibility in time analysis results: <ul style="list-style-type: none"> For both patients assessed over a period of 12 months measurements SD was 0.028 of the cosine correlation For the 1 border-line sample assessed over a period of 4 months measurements SD was 0.027 of the cosine correlation This latter sample was miss-classified 6 times (15%) | <p>The authors demonstrate for the first time that microarray technology can be used as a reliable diagnostic tool;</p> <p>The MammaPrint assay performed similarly to the original 70-genes signature</p> |

Evidence Table 8. Analytic validity, MammaPrint (continued)

| Study, year | Measure | Conclusions |
|-----------------------------|--|-------------|
| van't Veer ²¹ | Context: In this study the 70-gene signature was developed and no analytic validity data about the MammaPrint assay are reported. | |
| van de Vijver ²⁵ | Context: In this study the 70-gene signature was developed and no analytic validity data about the MammaPrint assay are reported. | |

RNA= ribonucleic acid; ANOVA= analysis of variance; Cy3= the green fluorescent dyes commonly used in two colors design microarray hybridization; Cy5= the red fluorescent dyes commonly used in two colors design microarray hybridization; LN= lymph node; SD= standard deviation;

Evidence Table 9. Clinical utility and validity, MammaPrint®

| Study, year | Measure | Conclusions |
|---------------------------|---|---|
| Buyse, 2006 ⁵⁹ | <p>Context: This study was designed to test the validation of the MammaPrint assay through the TRANSBIG consortium multicenter, multinational collaborative study involving 5 distinct institutions.</p> <p>End points: OS, RFS (any recurrence as event); TTM (time to metastases as first relapse)</p> <p>Methods:</p> <ul style="list-style-type: none"> • The Kaplan-Meier product limit estimator was used in survival analyses. • Hazard ratios were formulated to compare the high-risk group with the low-risk group. • Hazard ratios heterogeneity across institutions was evaluated by chi-square test with four degrees of freedom. • Gene signature hazard ratios were stratified by clinical risk and vice versa. • Follow-up duration impact on hazard ratios was analyzed by censoring all observations at increasing time points. • Sensitivity and specificity were estimated for distant metastases within 5 years and for death within 10 years. • Time-dependent receiver operating characteristic (ROC) curves were computed by using the tumor expression level for the 70-genes signature and the 10-year survival probability for the Adjuvant! Software. <p>Results:</p> <ul style="list-style-type: none"> • Kaplan-Maier analysis stratified by MammaPrint and Adjuvant! (% of patients with distant recurrence): <ul style="list-style-type: none"> Good(R>0.4), Adjuvant!Low: 52 patients OS(10years): 0.88 (0.74 to 0.95) Good(R>0.4)Adjuvant!High: 59 patients OS(10years): 0.89 (0.77 to 0.95) Poor(R<0.4)Adjuvant!Low: 28 patients OS(10years): 0.69 (0.45 to 0.84) Poor(R<0.4)Adjuvant!High: 163 patients OS(10years): 0.69 (0.61 to 0.76) | <p>MammaPrint is a better predictor of TTM than Age, Size, Grade, ER, Adjuvant!, NPI, St Gallen;</p> <p>St Gallen is a better predictor of DFS than MammaPrint</p> <p>MammaPrint is a better predictor for OS than Age, Size, Grade, ER, Adjuvant!, NPI, St Gallen;</p> <p>The signature remained a statistically significant prognostic factor for TTM and OS even after adjustment for various risk classifications methods based on clinicopathologic factors;</p> <p>The lack of statistical significance for DFS was explained by the fact that the signature was originally developed using TTM as the endpoint.</p> <p>Overall the 70-gene signature adds independent prognostic information to clinicopathologic risk assessment for node-negative untreated patients with early breast cancer;</p> <p>Clinical risk hazard ratios, adjusted for the gene signature were not significant, suggesting that most of their prognostic utility is subsumed by the gene signature;</p> |

Evidence Table 9. Clinical utility and validity, MammaPrint (continued)

| Study, year | Measure | Conclusions |
|------------------------|---|-------------|
| Buyse, 2006 (con't) | <ul style="list-style-type: none"> ● Hazard Ratios (unadjusted): <ul style="list-style-type: none"> MammaPrint: <ul style="list-style-type: none"> TTM=2.32 (95% CI = 1.35-4.00) DFS=1.50 (95% CI = 1.04-2.16) OS=2.79 (95% CI = 1.60-4.87) Adjuvant: <ul style="list-style-type: none"> TTM=1.68 (95% CI = 95% CI = 0.92 to 3.07) DFS=1.30 (95% CI = 95% CI = 0.86 to 1.95); OS=1.67 (95% CI = 95% CI =0.93 to 2.98) ● Hazard Ratios (adjusted): <ul style="list-style-type: none"> MammaPrint adjusted by St Gallen: <ul style="list-style-type: none"> TTM=2.15 (95% CI = 1.25 to 3.71) DFS=1.41 (95% CI = 0.97 to 2.06) OS=2.69 (95% CI = 1.53 to 4.73) MammaPrint adjusted by NPI: <ul style="list-style-type: none"> TTM=2.15 (95% CI =1.19 to 3.92) DFS=1.45 (95% CI = 0.97 to 2.16) OS=2.89 (95% CI = 1.58 to 5.29) MammaPrint adjusted by Adjuvant: <ul style="list-style-type: none"> TTM= 2.13 (95% CI = 1.19 to 3.82) DFS= 1.36 (95% CI = 0.91 to 2.03) OS= 2.63 (95% CI =1.45 to 4.79) Adjuvant Adjusted by MammaPrint: <ul style="list-style-type: none"> TTM=1.26 (95% CI = 0.66 to 2.40); DFS=1.03 (95% CI = 0.67 to 1.59) OS=1.08 (95% CI = 0.59 to1.99) ● Sensitivity for development of metastases within 5 years: <ul style="list-style-type: none"> Gene signature 0.90 (0.78 to 0.95) Adjuvant! software 0.87 (0.75 to 0.94) Nottingham Prognostic Index 0.79 (0.65 to 0.88) St Gallen criteria 0.96 (0.86 to 0.99) ● Specificity for development of metastases within 5 years: <ul style="list-style-type: none"> Gene signature 0.42 (0.36 to 0.48) Adjuvant! software 0.29 (0.24 to 0.35) Nottingham Prognostic Index 0.48 (0.42 to 0.54) St Gallen criteria 0.10 (0.07 to 0.14) | |

Evidence Table 9. Clinical utility and validity, MammaPrint (continued)

| Study, year | Measure | Conclusions |
|------------------------|--|-------------|
| Buyse, 2006 (con't) | <ul style="list-style-type: none"> ● Sensitivity for death within 10 years: Gene signature 0.84 (0.73 to 0.92) Adjuvant! software 0.82 (0.71 to 0.90) Nottingham Prognostic Index 0.70 (0.57 to 0.80) St Gallen criteria 0.95 (0.86 to 0.98) ● Specificity for death within 10 years: Gene signature 0.42 (0.36 to 0.48) Adjuvant! software 0.29 (0.23 to 0.35) Nottingham Prognostic Index 0.47 (0.41 to 0.53) St Gallen criteria 0.10 (0.07 to 0.14) ● ROC area under the curve for MammaPrint: TTM: 0.681 OS: 0.659 ● ROC area under the curve for Adjuvant: TTM: 0.648 OS: 0.576 ● Adjustments for clinical risk using different cutoff points produced only minor changes in the adjusted gene signature hazard ratios | |

Evidence Table 9. Clinical utility and validity, MammaPrint (continued)

| Study, year | Measure | Conclusions |
|------------------------------|---|---|
| Espinosa, 2005 ⁸⁰ | <p>Context: 70-genes signature partial validation using RT-PCR analysis: 60 out of the 70 genes of the signatures plus 4 additional genes. The study was conducted to reproduce the results obtained by microarray by an alternative method using 96 patients with full follow-up and frozen specimen from a single institution</p> <p>End points: Relapse-free survival (RFS), and overall survival (OS) measured by relapses dichotomization (regardless of the site of first relapse): low-risk if the patient survived 2 years or longer, or high risk otherwise</p> <p>Methods:</p> <ul style="list-style-type: none"> • The RT-PCR based 70-genes signature surrogate: was obtained as follows: <ul style="list-style-type: none"> Gene by gene qRT-PCR normalization, Gene expression categorization using the cutoff value of 0, Distance/similarity coefficient calculation for each patient, Similarity cutoff value optimization to yield the highest significance in log-rank tests for DFS • The Kaplan-Meier product limit method was used in survival analysis and significance was assessed by bilateral log-rank test • The Mann-Whitney test was used to compare patients' characteristics • The Fisher's exact test was used to compare categorical variables • Multivariate Cox proportional regression analysis was performed by adjusting by ER status, lymph node status, number of positive lymph nodes (less or more than 3) and gene profile <p>Results:</p> <ul style="list-style-type: none"> • Kaplan-Meier estimates for the Good prognosis profile: <ul style="list-style-type: none"> RFS after 70 months: 85% OS after 70 months: 97% Log-rank test P value = 0.03 • Kaplan-Meier estimates for the Bad prognosis profile: <ul style="list-style-type: none"> RFS after 70 months: 62% OS after 70 months: 72% Log-rank test P value = 0.002 • Results of multivariate Cox proportional regression analysis: <ul style="list-style-type: none"> OS by lymph node status, HR = 1.2; 95%CI = 1.09 to 1.36 OS by the 70-genes profile HR = 6.3; 95%CI = 1.28 to 31.07 DFS by lymph node number, HR = 1.13; 95%CI = 1.05 to 1.25 DFS by the 70-genes profile HR = 2.74, 95%CI = 1.13 to 6.61 | <p>qRT-PCR reproduced the results obtained with microarrays for the 70-genes prognostic profile in women with early-stage breast cancer</p> <p>Both RFS and OS were significantly longer in patients harboring a good-prognosis gene profile.</p> |

Evidence Table 9. Clinical utility and validity, MammaPrint (continued)

| Study, year | Measure | Conclusions |
|------------------------|---|--|
| Fan, 2006 ⁹ | <p>Context: The study is a comparison of classification agreement among 5 different gene expression based predictors (70-genes, Oncotype DX™, Wound-response, Intrinsic subtypes and two-gene ratio) using the 295 samples from the consecutive cohort originally used by van de Vijver 2002²⁵.</p> <p>Methods:</p> <ul style="list-style-type: none"> ● Cramer V statistics ● Kaplan-Meier survival analysis ● Multivariate Cox proportional hazards analysis, adjusting for age, tumor size, tumor grade, ER status, number of lymph node involved ● All analyses were performed using all the patients, as well as the ER positive patients subset (N=225) ● NB: coefficients of clinical predictors were allowed to vary between models <p>Results:</p> <ul style="list-style-type: none"> ● In multivariate Cox proportional hazards analysis all tests except the two-gene ratio were highly significant predictors of OS and DFS: 70 genes DFS HR = 3.4 (95%CI = 2.0-6.0), P value < 0.001 Two-gene ratio DFS HR = 0.91 (95%CI = 0.6-1.3), P value = 0.62 Oncotype OS HR = 4.3 (95%CI = 2.1-8.9), P value < 0.001 70 genes OS HR = 4.71 (95%CI = 2.02-11.0), P value < 0.001 Two-gene ratio OS HR = 1.00 (95%CI = 0.61-1.63), P value = 0.99 Oncotype DFS HR = 6.14 (95%CI = 1.84-20.4), P value = 0.003 ● ER status, tumor grade, tumor size, and lymph nodes were also significant predictors. ● Patients classification into dichotomized risks groups proved to be roughly similar, ranging from ~80% to ~40%: 70-genes and RS yielded a V= 0.60 70-genes and RS yielded an agreement of ~81% (239/295) | <p>Good but not perfect correlation between predictions, albeit surprising with different gene sets. Degree of prediction over and above “standard” clinical stratifiers is not clear – reclassification not done.</p> |

Evidence Table 9. Clinical utility and validity, MammaPrint (continued)

| Study, year | Measure | Conclusions |
|--------------------------|--|--|
| Glas, 2006 ⁵⁸ | <p>Context: MammaPrint assay development through reanalysis of patients from the van't Veer²¹ and van de Vijver²⁵ cohorts for which an RNA aliquot or the frozen specimen were available. A different reference RNA was used, as well as a different quantification method.</p> <p>Methods:</p> <ul style="list-style-type: none"> • 162 lymph node negative, young breast cancer patients (<55 years), who did not received adjuvant therapy. • All 78 patients from the van't Veer 2002²¹ cohort were available and used to reestablish the 70-genes expression profile, using the same procedures originally used and the same correlation to the signature (0.4). • 145 sample from the van de Vijver²⁵ cohort (61 also in van't Veer²¹) were used and a higher correlation was used for classification purposes (0.55). <p>Results:</p> <ul style="list-style-type: none"> • Comparison between MammaPrint and the original 70-genes signature using all the 78 patients from the van't Veer²¹ series for the risk of developing a metastasis as first relapse event within 5 years from surgery: MammaPrint OR = 13.95 (95%CI = 3.9-44) Original 70-genes signature OR = 15, 95%CI = 2.1 to 19) • 7 patients out of 78 were classified differently by MammaPrint with respect to the original 70-genes signature: <ul style="list-style-type: none"> 2 good prognosis patients were correctly classified by MammaPrint, but not by the original 70-genes signature 1 poor prognosis patient was correctly classified by MammaPrint, but not by the original 70-genes signature 2 poor prognosis patients were misclassified by MammaPrint, but not by the original 70-genes signature 2 good prognosis patients were misclassified by MammaPrint, but not by the original 70-genes signature • Comparison between MammaPrint and the original 70-genes signature using 145/151 LN negative patients from the van de Vijver²⁵ series for the risk of developing a metastasis as first relapse event within 5 years from surgery: HR for MammaPrint = 5.6 (95%CI = 2.4-7.3, p value = 0.0001) • The probability of remaining free of metastases at 5 years was not different between MammaPrint and the original 70-genes signature, with a p value = 0.89 in log-rank test • Similar results were obtained for OS | <p>The authors demonstrate for the first time that microarray technology can be used as a reliable diagnostic tool</p> <p>The MammaPrint assay performed similarly to the original 70-genes signature and is, therefore, an excellent tool to predict outcome of disease in breast cancer patients</p> |

Evidence Table 9. Clinical utility and validity, MammaPrint (continued)

| Study, year | Measure | Conclusions |
|-----------------------------------|--|--|
| van de Vijver, 2002 ²⁵ | <p>Context:70-genes signature partial validation using a gene expression array containing 25,000 features was used to evaluate the prognostic performance of the 70-genes signature in a consecutive series of 295 young women (< 53 years at diagnosis), with early stage (I-II), ER-positive or negative, node positive or negative breast cancer, who received or not adjuvant therapy (hormonal therapy, chemotherapy or both).</p> <p>End point: distant metastases as first relapse event; patients were censored on the date of the last follow-up visit, death from causes other than breast cancer, the recurrence of local or regional disease, or the development of a second primary cancer, including contralateral breast cancer and Overall Survival (OS) In this study, 61 of the 78 patients of the van't Veer²¹ cohort were included.</p> <p>Methods:</p> <ul style="list-style-type: none"> ● Kaplan-Meier analysis on the time from surgery to the first metastasis event or to censoring. ● Survival curves were compared by the log-rank test. ● Values are expressed as means ±SE, calculated according to the method of Tsiatis. ● All SEs were calculated with use of the sandwich estimator. ● Cox proportional-hazards regression analysis was used to investigate the association between the correlation coefficient (C1) and metastases by adjusting for other clinical variables (tumor grade, vascular invasion, and number of involved axillary's lymph-nodes: 0 versus 1 to 3 or 0 versus > 4). ● The linearity of the relation between the relative hazard ratio and the diameter of the tumor, age, and ER level of expression was tested with use of the Wald test for nonlinear components of restricted cubic splines. ● Hazard ratio proportionality was assessed by the method of Grambsch and Therneau. ● Difference between the relative hazard ratio before and after five years of follow-up with respect to the 70-genes prognosis signature was done by the Wald test. | <p>The 70-genes signatures is a better predictor of the risk of distant metastases than standard clinical predictors</p> |

Evidence Table 9. Clinical utility and validity, MammaPrint (continued)

| Study, year | Measure | Conclusions |
|-----------------------------|---|-------------|
| van de Vijver, 2002 (con't) | <p>Results:</p> <ul style="list-style-type: none"> ● The 70-genes gene expression signature: <ul style="list-style-type: none"> Proved to be associated with age, tumor grade, ER status (P value<0.001), and tumor size (P value=0.012) Was not associated with lymph-nodes, vascular invasion, surgery, chemotherapy, hormonal therapy ● OR for the various subgroups of patients in the consecutive series analyzed: <ul style="list-style-type: none"> 78 node negative patients used in van't Veer 2002²¹ <ul style="list-style-type: none"> 31 Poor signature with metastases 18 Poor signature without metastases 3 Good signature with metastases 26 Good signature with out metastases OR = 15, (95%CI = 3.3-56, p value < 0.001) 67 node-negative patients NOT used in van't Veer 2002: <ul style="list-style-type: none"> 11 Poor signature with metastases 23 Poor signature without metastases 1 Good signature with metastases 32 Good signature with out metastases OR = 15.3, (95%CI = 1.8-127, p value = 0.003) 180 node-positive and negative patients NOT used in van't Veer 2002: <ul style="list-style-type: none"> 39 Poor signature with metastases 65 Poor signature without metastases 3 Good signature with metastases 73 Good signature with out metastases OR = 14.6, (95%CI = 4.3-50, p value < 0.0001) ● Probability of metastases at 5yrs for all the patients: <ul style="list-style-type: none"> 115 Good signature patients: 94.7±2.1SE 180 Poor signature patients: 60.5±3.8SE ● Probability of metastases at 10 yrs for all the patients <ul style="list-style-type: none"> 115 Good signature patients: 85.2±4.3SE 180 Poor signature patients: 50.6±4.5SE ● Probability of metastases at 5 yrs for node negative patients: <ul style="list-style-type: none"> 93 Good signature patients: 4±3.2SE 91 Poor signature patients: 56.2±5.5SE ● Probability of metastases at 10 yrs for node negative patients: <ul style="list-style-type: none"> 60 Good signature patients: 86.8±4.8SE 91 Poor signature patients: 44.1±6.3SE ● Probability of metastases at 5 yrs for node positive patients <ul style="list-style-type: none"> 55 Good signature patients: 95.2±2.6SE 89Poor signature patients: 66.3±5.2SE | |

Evidence Table 9. Clinical utility and validity, MammaPrint (continued)

| Study, year | Measure | Conclusions |
|-----------------------------|--|-------------|
| van de Vijver, 2002 (con't) | <ul style="list-style-type: none"> ● Probability of metastases at 10 yrs for node positive patients <ul style="list-style-type: none"> 55 Good signature patients: 82.7±7.8SE 89 Poor signature patients: 56.7±6.4SE ● OS at 5 yrs for all the patients: <ul style="list-style-type: none"> 15 Good patients 97.4±1.5SE 180 Poor patients 74.1±3.3SE ● OS at 10 yrs for all the patients: <ul style="list-style-type: none"> 115 Good patients 94.5±2.6SE 180 Poor patients 54.6±4.4SE ● OS at 5 yrs for node-negative patients: <ul style="list-style-type: none"> 60 Good patients 96.7±2.3SE 91 Poor patients 71.5±4.8SE ● OS at 10 yrs for node-negative patients: <ul style="list-style-type: none"> 60 Good patients 96.7±2.3SE 91 Poor patients 49.6±6.1SE ● OS at 5 yrs for node-positive patients: <ul style="list-style-type: none"> 55 Good patients 98.2±1.8SE 89 Poor patients 76.5±4.6SE ● OS at 10 yrs for node-positive patients: <ul style="list-style-type: none"> 55 Good patients 92.0±4.8SE 89 Poor patients 59.5±6.3SE ● HAZARD RATIO for metastases as first event over the entire follow-up period: <ul style="list-style-type: none"> All patients, HR = 5.1, (95%CI = 2.9-9.0, p value < 0.001) 151 node negative patients, HR = 5.5, (95%CI = 2.5-12.2, p value < 0.001) 144 node positive patients, HR = 4.5, (95%CI = 2-10.2, p value < 0.001) ● HAZARD RATIO for OS over the entire follow-up period: <ul style="list-style-type: none"> All patients = 8.6, (95%CI = 4-19, p value < 0.001) ● In multivariate analysis the following factors resulted to be independent predictors of risk of metastases as first event within 5 years: <ul style="list-style-type: none"> 70-genes signature, HR = 4.6, 95%CI = 2.3-9.2, p value < 0.001 Tumor size, HR = 1.56, 95%CI = 1.22-2.0, p value < 0.001 Chemotherapy, HR = 0.37, 95%CI = 0.2-0.66, p value < 0.001 The other covariates used (age, grade, node status, vascular invasion, ER status, surgery, and hormonal therapy) were not associated with the risk of metastasis | |

Evidence Table 9. Clinical utility and validity, MammaPrint (continued)

| Study, year | Measure | Conclusions |
|-----------------------------|---|-------------|
| van de Vijver, 2002 (con't) | <ul style="list-style-type: none"> ● Reclassification and comparison with St. Gallen and NIH criteria: The 70-genes signature classified more node negative patients in the Good prognosis group than both the St. Gallen and NIH criteria (40%, 15% and 7% respectively) 43 St. Gallen High Risk group patients were re-classified in the Good signature group 86 St. Gallen High Risk group patients were classified in the Poor signature group KM analysis showed that the 70-genes signature groups defined within the St. Gallen High Risk group are different, with a p value < 0.001 17 St. Gallen Low Risk group patients were classified in the Good signature group 5 St. Gallen Low Risk group patients were classified in the Poor signature group KM analysis showed that the 70-genes signature groups are PROBABLY different within the St. Gallen Low Risk group, with a p value < 0.11 53 NIH High Risk patients were classified in the Good signature group 87 NIH High Risk patients were classified in the Poor signature group KM analysis showed that the 70-genes signature groups are different within the NIH High Risk group, with a p value < 0.001 7 NIH-Low Risk group patients were classified in the Good signature group 4 NIH-Low Risk group patients were classified in the Poor signature group KM analysis showed that the 70-genes signature groups are different within NIH Low Risk group, with a p value < 0.05; | |

Evidence Table 9. Clinical utility and validity, MammaPrint (continued)

| Study, year | Measure | Conclusions |
|--------------------------------|--|---|
| van't Veer, 2002 ²¹ | <p>Context: 70-genes gene expression signature development using a gene expression array containing 25,000 features was used to select genes associated with metastases free survival at 5 years from the surgery.</p> <p>End point: distant metastases as first relapse event; patients were censored on the date of the last follow-up visit, death from causes other than breast cancer, the recurrence of local or regional disease, or the development of a second primary cancer, including contralateral breast cancer.</p> <p>Methods:</p> <ul style="list-style-type: none"> ● Kaplan-Meier analysis on the time from surgery to the first metastasis event or to censoring. ● Survival curves were compared by the log-rank test. ● Multivariate proportional-hazards regression analysis was used to investigate the association between the correlation coefficient and the time to metastases by adjusting for other clinical variables. ● In the multivariate analysis the following predictors were used: tumor grade, ER status, PR status, tumor size, patient age, and angioinvasion. ● Values were expressed as means ±SE, calculated according to the method of Tsiatis. ● All SEs were calculated with use of the sandwich estimator. <p>Results:</p> <ul style="list-style-type: none"> ● Optimized 70-genes signature performance in the training set (78 patients): 65/78 correct predictions (patients developed or not metastases within 5 years) 13 misclassified patients: 5 patients with poor prognosis were classified in the 70-gene Good prognosis group 8 patients with good prognosis were classified in the 70-gene Poor prognosis group ● Optimized 70-genes signature performance in the validation set (19 patients): 17/19 correct predictions (patients developed or not metastases within 5 years) 2 misclassified patients: 1 patients with poor prognosis was classified in the 70-gene Good prognosis group 1 patients with good prognosis in the 70-gene Poor prognosis group ● ODDS RATIO to develop metastases within 5 years for a patient in the 70-genes Poor prognosis group: OR=28, (95%CI=7-107), P value = 0.00000001 Leave-one-out cross-validated OR =15, (95%CI=4-56), P value=0.0000041 ● In univariate analysis the 'leave-one-out' cross-validated 70 genes signature proved to be a better predictor of recurrence than grade, size(>2cm), age (<40yrs), ER negative status, and angioinvasion | <p>The 70-genes signatures is a better predictor of the risk of distant metastases than standard clinical predictors;</p> |

Evidence Table 9. Clinical utility and validity, MammaPrint (continued)

| Study, year | Measure | Conclusions |
|--------------------------|--|-------------|
| van't Veer, 2002 (con't) | <ul style="list-style-type: none"> ● In multivariate analysis the 70-genes signature showed to be an independent predictor of metastases within 5 years (logistic regression OR = 18, (95%CI=3.3-94, P value = 0.00014) ● Comparison of the 70 genes signatures with St. Gallen and NIH criteria for risk stratification: <ul style="list-style-type: none"> St. Gallen: <ul style="list-style-type: none"> 64/78 patients were defined as at high risk of metastases (33/34 of the patients who did develop metastases; 31/44 who did not develop metastases) NIH: <ul style="list-style-type: none"> 72/78 patients were defined as at high risk of metastases (32/34 of the patients who did develop metastases; 40/44 who did not develop metastases) 70-genes signature: <ul style="list-style-type: none"> 43/78 patients were defined as at high risk of metastases (31/34 of the patients who did develop metastases; 18/44 who did not develop metastases); | |

TRANSBIG= Breast International Group (BIG) network for improved treatment tailoring; OS= overall survival; RFS = relapse free survival; TTM = time to distant metastases; ROC= receiver operating characteristic; ER= estrogen receptor; NPI= Nottingham Prognostic Index; CI= confidence interval; RT-PCR= real time polymerase chain reaction; qRT-PCR= quantitative real time polymerase chain reaction; HR= hazard ratio; RNA= ribonucleic acid; OR= odds ratio; SE= standard error; NIH= National Institutes of Health; Kaplan Meier= Kaplan Meier; PR= partial relapse.

Evidence Table 10. Study design, H/I ratio

| Study, Year | Country | Study (data collection) period | Study Type | Population size, N | Blinded (Y/N) | Study purpose |
|------------------------------|----------------|---------------------------------------|-------------------|-----------------------------------|----------------------|---|
| Fan, 2006 ⁷⁹ | Netherlands | NR | retrospective | Poor prognosis, 180 | NA | Validation of signature |
| | | | | Good prognosis, 115 | | |
| Goetz, 2006 ⁶² | USA | NR | RCT | With expression ratio data, 206 | NR | Validation of signature |
| | | | | Without expression ratio data, 50 | | |
| Jansen, 2007 ⁷² | Netherlands | 1979-1996 | retrospective | LN-/ER-, 185 | NR | Validation of signature |
| | | | | LN-/ER+, 468 | | |
| | | | | LN+/ER-/NO-, 41 | | |
| | | | | LN+/ER+/NO-, 151 | | |
| | | | | LN+/ER uk/YES adjuvant, 407 | | |
| Jerevall, 2007 ⁶³ | Sweden | 1982 - 1995 | retrospective | Post-menopausal, 264 | NR | Validation of signature |
| Ma, 2004 ⁶⁴ | USA | 1987-1997 | retrospective | Frozen, recurrence, 28 | NR | Validation and development of signature |
| Ma, 2006 ⁶¹ | USA | 2001 | prospective | Tomoxifen treated, 286 | NR | Validation and development of assay |
| | | | | Untreated, 566 | | |
| | | | | Frozen, non-recurrence, 32 | | |
| | | | | FFPE, recurrence, 10 | | |
| | | | | FFPE, non-recurrence, 10 | | |
| Reid, 2005 ⁶⁹ | USA | NR | prospective | Tamoxifen treated 58 | NR | Validation of signature |
| | | | | Pre-menopausal, 93 | | |

NR = not reported; NA = not applicable; LN = lymph node; ER = estrogen receptor, NO = node; uk = unknown; FFPE =formalin fixed paraffin embedded.

Evidence Table 11. Study population characteristics, H/I ratio (continued)

| Study, Year | Intervention | General Characteristics | Diagnosis (es)* | Treatments and Outcomes |
|---------------------------|---|---|--|--|
| Goetz, 2006 ⁶² | Tamoxifen 20mg/day for 5 years | Age, median (range): 68 (42-84) Menopausal status: 100 % Race (%): W (92) Exclusion criteria: NR | H Ductal; 86% H Lobular; 10% H Other; 4% TS <3; 76% TS ≥3; 24% LN+ nodes 0, 63% LN+ nodes 1-3, 26% LN+ nodes 4-9, 7% LN+ nodes ≥10, 4% ER 10-49 fmol; 20% ER ≥50 fmol; 69% ER +; 11% HER2 0; 11% HER2 1; 36% HER2 2; 34% HER2 Unknown; <1% Nottingham TG 1, 26% Nottingham TG 2, 55% Nottingham TG 3, 18% Nottingham TG Unknown, <1% Diagnostic method: RT-PCR | Treatment: s: Mastectomy; 83% s: Breast Conservation; 17% Outcome NR |
| | Tamoxifen 20mg/day for 5 years (women without expression ratio data-all treated with oral tamoxifen) | Age, median (range): 64 (48-87) Menopausal status: 100 % Race (%): W (91) Exclusion criteria: NR | H Not Determined TS <3; 90% TS ≥3; 10% LN+ nodes 0, 62% LN+ nodes 1-3, 15% LN+ nodes 4-9, 71% LN+ nodes ≥10, 6% ER 10-49 fmol; 20% ER ≥50 fmol; 56% ER+; 24% HER2 Not determined Nottingham TG; Not determined Diagnostic method: RT-PCR | Treatment: s: Mastectomy; 74% s: Breast Conservation; 26% Outcome: NR |

Evidence Table 11. Study population characteristics, H/I ratio (continued)

| Study, Year | Intervention | General Characteristics | Diagnosis (es)* | Treatments and Outcomes |
|----------------------------|---|---|--|--|
| Jansen, 2007 ⁷² | 1,683 total female patients diagnosed with breast cancer from 1979-1996, clinical information available for 1,252 primary tumor specimens, subgroups (from figure1) | Age: NR Menopausal status: NR Race: NR Exclusion criteria: distant spread at or within the first month of surgery | Premenopausal: LN+ >3Nodes; 330 (27%) LN+ 1-3Nodes; 269 (21%) LN-; 653 (52%) TS <2cm; 392 (31%) TS 2-5cm; 716 (57%) TS >5cm/pT4; 144 (12%) ER High; 917 (73%) ER Low; 335 (27%) PR High; 764 (61%) PR Low; 417 (33%) | Treatment received by 406 patientst: s: breast-conserving lumpectomy, 543 (43) mastectomy, 709 (57) h: 177 patients (44) a: 214 (53) c: 15 (3) Follow-up: 72 months, all patients (2-248 months) 91 months, 692 patients alive (3 to 248 months) 46 months , 560 deaths (2 to 205 months) Outcome: Disease recurrence occurred in 692 (55%) of 1,252 patients. 846 patients (68%) did not receive neo-adjuvant systemic therapy ● 381 patients (45%) experienced a relapse, of which 285 (75%) were treated with hormonal therapy of which;193 patients (68%) received first-line tamoxifen monotherapy, of which: Therapy failure occurred in 73 patients (38%), of which:61 patients (84%) had progressive disease; 12 patients (16%) showed stable disease < 6 months; ● 381 patients (45%) experienced a relapse, 193 patients (68%) received first-line tamoxifen monotherapy, of which: 120 (62%) of these 193 |
| | LN-/ER- = 185 (15%) | Age, n (%): < 40, 166 (13) Menopausal status: pre, n=537 (43%) Race: NR Exclusion criteria: missing values for lymph node status, ER protein status, and HOXB13 and IL17BR | Diagnostic method: RT-PCR note1: The cutoff point to classify primary breast tumors as ER and/or progesterone receptor (PgR) -positive was 10 fmol/mg cytosolic protein. | |
| | LN-/ER+ = 468 (37%) | Age, n (%): 40-55,: 465 (37) Menopausal status: post, n=715 (57%) Race: NR Exclusion criteria: 30% epithelial tumor cells in their tumor specimens | note2: Follow-up, tumor staging, and response to therapy was defined by standard International Union Against Cancer (Geneva, Switzerland) classification criteria | |
| | LN+/ER-/LN-adjuvant = 41 (3%) | Age, n (%): 56-70, 409 (33) Menopausal status: NR Race: NR Exclusion criteria: specimens of poor RNA quality | | |
| | LN+/ER+/LN-adjuvant = 151 (12%) | Age, n (%): >70, 212 (17) Menopausal status: NR Race: NR Exclusion criteria: NR | | |

Evidence Table 11. Study population characteristics, H/I ratio (continued)

| Study, Year | Intervention | General Characteristics | Diagnosis (es)* | Treatments and Outcomes |
|------------------------------|---|--|--|--|
| Jansen, 2007 (con't) | LN+/ER uk/ YES adjuvant = 407 (33%) | Age: NR Menopausal status: NR Race: NR Exclusion criteria: NR | | patients experienced a clinical benefit from first-line tamoxifen, of which: disease-free survival (DFS) in patients with primary breast cancer (N = 619) and progression free; survival (PFS) in patients with recurrent breast cancer treated with first-line tamoxifen monotherapy (N = 193). |
| Jerevall, 2007 ⁶³ | <p>Postmenopausal patients: randomized clinical trial, comparing 2 years (163 patients, 62%) and 5 years (101 patients, 38%) of adjuvant tamoxifen treatment</p> <p>Premenopausal patients: clinical trial comparing tamoxifen treatment versus no treatment, only untreated patients were used</p> | Age: NR Menopausal status: NR Race: NR Exclusion criteria: NR | <p>Postmenopausal: LN+; 73 (28%) LN-; 184 (72%) LN Unknown; 7 (?%) TS <2cm; 77 (29%) TS ≥2cm; 187 (71%) ER +; 195 (74%) ER -; 69 (26%) PR +; 142 (54%) PR -; 120 (46%)</p> <p>Premenopausal: LN+; 63 (28%) LN-; 67 (72%) TS <2cm; 24 (26%) TS ≥2cm; 69 (74%) ER +; 59 (63%) ER -; 34 (37%) PR +; 60 (65%) PR -; 33 (35%) Nottingham TG 4, 4 (4.9%) Nottingham TG 2, 38 (47%) Nottingham TG 3, 39 (48%) Nottingham TG Unknown, 12 (?%)</p> <p>Diagnostic method: RT-PCR</p> | Treatment: s: Modified radical mastectomy s: breast conserving surgery with axillary lymph node dissection ; radiotherapy (50 Gy/25 fractions) to lymph node-positive patients ; |

Evidence Table 11. Study population characteristics, H/I ratio (continued)

| Study, Year | Intervention | General Characteristics | Diagnosis (es)* | Treatments and Outcomes |
|------------------------|---|--|--|---|
| Ma, 2004 ⁶⁴ | Frozen, recurrence, n=28 | Age, mean y (range): 65.1 (48-84) Menopausal status: NR Race: NR Exclusion criteria: NR | TS mean, 2.7 TS range, 0.9-4.7 LN# 0, 13 LN# 1-3, 6 LN# >3, 6 TG 1, 2 TG 2, 15 TG 3, 11 ER+, 27 PR+, 23 Diagnostic method: microarray and RT-PCR | Treatment: NR Outcome: DFS: mean months (range): 54.8 (5-137) |
| | Frozen, non recurrence, n=32 | Age, mean y (range): 69.1 (54-85) Menopausal status: NR Race: NR Exclusion criteria: NR | TS mean, 2.1 TS range, 0.8-5.5 LN# 0, 15 LN# 1-3, 11 LN# >3, 2 TG 1, 1 TG 2, 24 TG3, 7 ER+, 32 PR+, 27 Diagnostic method: microarray and RT-PCR | Treatment: NR Outcome: DFS: mean months (range): 115.6 (61-169) |
| | Formalin-fixed paraffin embedded (FFPE), recurrence, n=10 | Age, mean y (range): 65.5 (54-93) Menopausal status: NR Race: NR Exclusion criteria: NR | TS mean, 1.9 TS range, 1.1-4 LN# 0, 8 LN# 1-3, 1 LN# >3, 0 TG 1, 1 TG 2, 6 TG3, 3 ER+, 10 PR+, 8 Diagnostic method: RT-PCR | Treatment: NR Outcome: DFS: mean months (range): 51.4 (15-117) |

Evidence Table 11. Study population characteristics, H/I ratio (continued)

| Study, Year | Intervention | General Characteristics | Diagnosis (es)* | Treatments and Outcomes |
|---------------------|----------------------------|---|---|---|
| Ma, 2004 (con't) | FFPE, non recurrence, n=10 | Age, mean y (range): 65.2 (57-74) Menopausal status: NR Race: NR Exclusion criteria: NR | TS mean, 1.7 TS range, 0.8-4 LN# 0, 10 LN# 1-3, 0 LN# >3, 0 TG 1, 1 TG 2, 8 TG3, 1 ER+, 10 PR+, 10 Diagnostic method: RT-PCR | Treatment: NR Outcome: DFS: mean months (range): 95.8 (25-123) |
| Ma, 2006 61 | All samples, n=852 | Age, n (%): <50, 155 (18) >50, 695 (82) UK, 2 (0.2) Menopausal status: NR Race: NR Exclusion criteria: NR | TS<2, 309 (36) TS 2-5, 460 (54) TS >5, 75 (9) TS UK, 8 (1) LN# =0, 613 (72) LN#=1-3, 134 (16) LN#=4+, 105 (12) ER+, 625 (73) ER-, 214 (25) ER UK, 13 (2) HER2-CISH 0-3, 589 (69) HER2-CISH 4-12, 133 (16) HER2-CISH UK, 130 (15) PR+, 461 (54) PR-, 382 (45) PR UK, 9 (1) SP<6, 215 (25) SP 6-10, 213 (25) SP>10, 287 (34) SP UK, 137 (16) PL dip, 281 (33) PI An, 465 (55) PI UK, 106 (12) | Treatment: a: With or without adjuvant tamoxifen monotherapy s: Mastectomy or lumpectomy plus axillary dissection r: With or without postoperative radiation therapy Outcome: RFS: mean =70 (0-285) RS: 266 (31) D: cancer: 206 (24) D: other: 155 (18) |

Evidence Table 11. Study population characteristics, H/I ratio (continued)

| Study, Year | Intervention | General Characteristics | Diagnosis (es)* | Treatments and Outcomes |
|------------------|--------------------------|---|--|--|
| Ma, 2006 (con't) | Tamoxifen treated, n=286 | Age, n (%): <50, 26 (9) >50, 260 (91) UK, 0 Menopausal status: NR Race: NR Exclusion criteria: NR | TS<2, 102 (36) TS 2-5, 157 (55) TS >5, 26 (9) TS UK, 1 (0.1) LN# =0, 138 (40) LN#=1-3, 79 (28) LN#=4+, 69 (24) ER+, 255 (89) ER-, 29 (10) ER UK, 2 (1) HER2-CISH 0-3, 204 (71) HER2-CISH 4-12, 42 (15) HER2-CISH UK, 40 (14) PR+, 188 (66) PR-, 97 (34) PR UK, 1 (0.1) SP<6, 69 (24) SP 6-10, 89 (31) SP>10, 89 (31) SP UK, 39 (14) PL dip, 95 (33) PI An, 166 (58) PI UK, 25 (9) Diagnostic method: RT-PCR | Treatment: a: With or without adjuvant tamoxifen monotherapy s: Mastectomy or lumpectomy plus axillary dissection r: With or without postoperative radiation therapy Outcome: RFS: mean =68 (0-231) RS: 83 (29) D: cancer: 69 (24) D: other: 49 (17) |

Evidence Table 11. Study population characteristics, H/I ratio (continued)

| Study, Year | Intervention | General Characteristics | Diagnosis (es)* | Treatments and Outcomes |
|---------------------|------------------|---|---|---|
| Ma, 2006 (con't) | Untreated, n=566 | Age, n (%): <50, 129 (23) >50, 435 (77) UK, 2 (0.2) Menopausal status: NR Race: NR Exclusion criteria: NR | TS<2, 207 (37) TS 2-5, 303 (54) TS >5, 49 (9) TS UK, 7 (1) LN# =0, 475 (84) LN#=1-3, 66 (10) LN#=4+, 36 (6) ER+, 370 (65) ER-, 185 (33) ER UK, 11 (2) HER2-CISH 0-3, 385 (68) HER2-CISH 4-12, 91 (16) HER2-CISH UK, 90 (16) PR+, 273 (48) PR-, 285 (50) PR UK, 8 (1) SP<6, 146 (26) SP 6-10, 124 (22) SP>10, 198 (35) SP UK, 98 (17) PL dip, 186 (33) PI An, 299 (53) PI UK, 81 (14) Diagnostic method: RT-PCR | Treatment: a: With or without adjuvant tamoxifen monotherapy s: Mastectomy or lumpectomy plus axillary dissection r: With or without postoperative radiation therapy Outcome: RFS: mean =71 (0-285) RS: 183 (32) D: cancer: 137 (24) D: other: 106 (17) |

Evidence Table 11. Study population characteristics, H/I ratio (continued)

| Study, Year | Intervention | General Characteristics | Diagnosis (es)* | Treatments and Outcomes |
|--------------------------|---|--|--|------------------------------|
| Reid, 2005 ⁶⁹ | 58 patients treated with Tamoxifen 20mg/day for 5 years | Age, n (%): ≥50yrs, 4 (6.9) >50yrs, 54 (93.1) Menopausal status: NR Race: NR Exclusion criteria: NR | H Ductal; 48/58 (83.0%) H Lobular; 6/58 (10.2%) H Ductal + Lobular; 2/58 (3.4%) H Other; 2/58 (3.4%) TS ≤2cm; 22/58 (37.9%) TS >2cm; 36/58 (62.1%) LN# 0; 13/58 (22.5%) LN# 1-3; 31/58 (53.5%) LN# ≥4; 14/58 (24.0%) TG 1; 3/58 (5.0%) TG 2; 28/58 (48.3%) TG 3; 12/58 (20.7%) TG Not Assessed; 15/58 (26.0%) HER2 -; 46/58 (77.6%) HER2 +; 12/58 (20.7%) PR -; 12/58 (20.7%) PR +; 46/58 (79.3%) Diagnostic method: RT-PCR | Treatment: NR Outcome: NR |

* Diagnoses: H = histology/histologic type; TS = tumor size/diameter; LN# = # of Lymph nodes at primary diagnosis; LN = lymph node status; TG = tumor grade; ER = estrogen receptor status; HER = HER neu status; PR = Progesterone receptor status; SP = S-phase fraction; PI = ploidy; VI = vascular invasion; PTS = site of primary tumor; HR = hormonal status, ER or PR or both; DR = distant recurrence

† Treatments: H = hormonal; A = adjuvant; S = surgical; R = radiation; c = combination

‡ Treatment outcomes: OS = overall survival; RFS = relapse-free survival (first disease relapse, irrespective of site); CR = complete response; PR = partial response; RS: recurrence score; SD = stable disease; PD = progressive disease; PCR = pathologic complete response; D = death

§ Data not reported in article, referred directly to van de Vijver, 2002²⁵.

|| all patients were analyzed, however the following groups are of relevance: (LN-/ER+) and (LN+/ER+/NO-adjuvant); and 116 from (LN-/ER+) and 77 from (LN+/ER+/NO-adjuvant), total of 193 (15%), who received first-line tamoxifen treatment

NR= not reported; RT-PCR= reverse transcriptase polymerase chain reaction; DFS= disease free survival; FFPE= formalin-fixed paraffin embedded

Evidence Table 12. Analytic validity, H/I ratio

| Study, year | Measure | Conclusions |
|-------------------------------|---|-------------|
| Goetz, 2006 ⁶² | <p>Context: Validation of the H/I ratio signature using patients from the randomized NCCTG 89-30-52 trial on tamoxifen treatment. Out of the 256 eligible patients, FFPE were available for 227 patients</p> <p>Methods:</p> <ul style="list-style-type: none"> ● LMC was performed prior total RNA preparation ● RT-PCR expression values for each gene were normalized using a standard curve obtained analyzing the human universal total RNA (Stratagen, La Jolla, CA) ● No reference genes were used <p>Results:</p> <ul style="list-style-type: none"> ● 206 patients out of 256 successfully analyzed, for the following reasons: No specimen available: 29/256 patients (11%) Tumor content sufficient: 211/227 patients (93%) Failed RT-PCR: 5/211 patients (2.4%) Successful assays: 206/211 patients (97.6%) | |
| Jansen, 2007 ⁷² | <p>Context: Validation of the H/I ratio signature by measuring expression levels normalized to a different set of control genes respect to MA et al 2006⁶¹ using fresh frozen samples. From a population of 1693, 1,252 were considered eligible and were successfully evaluated</p> <p>Methods:</p> <ul style="list-style-type: none"> ● RT-PCR was use to measure the expression levels of HOXB13 and IL17BR ● Two pairs of primers-probes were used to evaluated the IL17BR gene, aligning to the 3' and the 5' end of the genes and corresponding to the region of the transcripts respectively assayed by Ma et al 2004⁶⁴ and Reid et al 2005⁶⁹ ● Three reference genes different from those used by Ma et 2006⁶¹ were used (porphobilinogen deaminase, hypoxanthine-guanine phospho-ribosyltransferase, and beta-2-microglobulin) <p>Results: In this study analytic validity data were not reported</p> | |

Evidence Table 12. Analytic validity, H/I ratio (continued)

| Study, year | Measure | Conclusions |
|------------------------------|---|-------------|
| Jerevall, 2007 ⁶³ | <p>Context: Validation of the H/I ratio signature. Expression levels were normalized to b-actin using . fresh frozen samples. Patients were collected from two distinct institutions; of 373 tumor samples analyzed, RNA expression data were obtained from 357 tumors</p> <p>Methods:</p> <ul style="list-style-type: none"> ● RT-PCR reactions in the two institutions were performed using the same sets of primers/probes and two distinct instruments ● In each reaction serial dilutions of a standard sample were used to construct a standard curve used to quantify gene expression prior normalization ● The reproducibility between the two institution was assessed by Pearson’s correlation <p>Results:</p> <ul style="list-style-type: none"> ● 357 patients out of 373 successfully analyzed, for the following reasons: Insufficient RNA yield: 16/373 patients (4.3%) Overall success rate: 357/373 patients (95.7%%) ● Reproducibility between institutions, Pearson’s correlations, 10 patients: HOXB13:b-actin, $r = 0.96$, $P < 0.001$ IL17BR:b-actin, $r = 0.87$, $P = 0.002$ HOXB13:IL17BR, $r = P < 0.001$ ● Median values for all the samples analyzed at the two centers: HOXB13:b-actin, 0.086 vs. 0.081 IL17BR:b-actin, 1.4 vs. 1.3) HOXB13:IL17BR, 0.074 vs. 0.055 | |
| Ma, 2004 ⁶⁴ | <p>Context: This study used the H/I ratio signature for the development and preliminary validation of the of the two-gene ratio signature. The training set was analyzed by microarray (n=60) and by RT-PCR (n=59)</p> <p>Methods:</p> <ul style="list-style-type: none"> ● Comparison of microarray based and RT-PCR based ratio by Pearson’s correlation on 59/60 patients <p>Results:</p> <ul style="list-style-type: none"> ● Correlations between array and RT-PCR: HOXB13, $r = 0.83$ IL17BR , $r = 0.93$ HOXB13/IL17BR, $r = 0.83$ | |

Evidence Table 12. Analytic validity, H/I ratio (continued)

| Study, year | Measure | Conclusions |
|------------------------|---|--|
| Ma, 2006 ⁶¹ | <p>Context: Development of the HOXB13:IL17BR two-gene index. Out of the 1,002 eligible The FFPE tissue microarrays were 4 years old, and obtained from specimens originally stored as fresh frozen tissues</p> <p>Methods:</p> <ul style="list-style-type: none"> ● IHC for ER and PR was performed^{90,91} ● IHC Allred⁹² scores of 3 to 8 were considered positive for ER or PR⁹³ ● Concordance between RT-PCR and IHC results by k statistics ● Since both ER and PR mRNA RT-PCR measurements were bimodal; the following midpoints were used as cutoffs: 2.5 CT for ER 5.9 for PR <p>Results:</p> <ul style="list-style-type: none"> ● The quality of these tissue microarrays FFPE specimens was comparable with those without the intervening snap-freeze step (data not shown) ● 852 patients out of 1002 successfully analyzed, for the following reasons: Tumor content < 10%: 132/1002 patients (13.2%) Poor RNA yield: 18/870 patients (2%) Successful assays: 852/870 (98%) ● Agreement between IHC vs RT-PCR: ER, 91% concordance, k = 0.83, P value = .0001 PR, 85% concordance, k = 0.70, P value = .0001 | <p>According to the authors these results confirmed the significant correlations between mRNA and protein levels for ER and PR and provided validation of the FFPE gene expression assay platform.</p> |

NCCTG= North Central Cancer Treatment Group; FFPE = formalin fixed paraffin embedded; LMC= laser micro-dissection; RNA= ribonucleic acid; RT-PCR = reverse transcriptase polymerase chain reaction; IHC = immunohistochemistry; ER = estrogen receptor; PR = progesterone receptor; mRNA= messenger ribonucleic acid.

Evidence Table 13. Clinical utility and validity, H/I ratio

| Study, year | Measure | Conclusions |
|------------------------|--|--|
| Fan, 2006 ⁹ | <p>Context: The study is a comparison of classification agreement among 5 different gene expression based predictors (70-genes, Oncotype DX, Wound-response, Intrinsic subtypes and two-gene ratio) using the 295 samples from the consecutive cohort originally used by van de Vijver 2002[reference].</p> <p>Methods:</p> <ul style="list-style-type: none"> ● Cramer V statistics ● Kaplan-Meier survival analysis ● Multivariate Cox proportional hazards analysis, adjusting for age, tumor size, tumor grade, ER status, number of lymph node involved ● All analyses were performed using all the patients, as well as the ER positive patients subset (N=225) ● NB: coefficients of clinical predictors were allowed to vary between models <p>Results:</p> <ul style="list-style-type: none"> ● In multivariate Cox proportional hazards analysis all tests except the two-gene ratio were highly significant predictors of OS and DFS: <ul style="list-style-type: none"> 70 genes DFS HR = 3.4 (95%CI = 2.0-6.0), P value < 0.001 Two-gene ratio DFS HR = 0.91 (95%CI = 0.6-1.3), P value = 0.62 Oncotype OS HR = 4.3 (95%CI = 2.1-8.9), P value < 0.001 70 genes OS HR = 4.71 (95%CI = 2.02-11.0), P value < 0.001 Two-gene ratio OS HR = 1.00 (95%CI = 0.61-1.63), P value = 0.99 Oncotype DFS HR = 6.14 (95%CI = 1.84-20.4), P value = 0.003 ● ER status, tumor grade, tumor size, and lymph nodes were also significant predictors. ● Patients classification into dichotomized risks groups proved to be roughly similar, ranging from ~80% to ~40%: <ul style="list-style-type: none"> 70-genes and RS yielded a V= 0.60 70-genes and RS yielded an agreement of ~81% (239/295) | <p>Good but not perfect correlation between predictions, albeit surprising with different gene sets. Degree of prediction over and above “standard” clinical stratifiers is not clear – reclassification not done.</p> |

Evidence Table 13. Clinical utility and validity, H/I ratio (continued)

| Study, year | Measure | Conclusions |
|------------------------------|---|---|
| Goetz, 2006 ⁶² | <p>Context: Validation of the H/I ratio signature using patients from a randomized trial on tamoxifen treatment.</p> <p>End points:</p> <ul style="list-style-type: none"> ● RFS (time from randomization to any event of recurrence, or contralateral breast cancer or death), DFS (time from randomization to any event of recurrence, or contralateral breast cancer, or other cancer, or death), and OS (time from randomization to death) ● RT-PCR expression values for each gene were normalized using a standard curve obtained analyzing the human universal total RNA (Stratagen, La Jolla, CA); No reference genes were used ● Specific cut-off points were obtained in the selected population analyzed: none negative (n=130) and node-positive (n=86) patients <p>Methods:</p> <ul style="list-style-type: none"> ● Cut-off points estimation for the HOXB13:IL-17BR ratio were obtained by minimizing the P value in log-rank tests ● Multiple testing corrected P values from the log-rank test were obtained by the Lausen and Schumacher method, modify by Altman ● Log-rank test and univariate Cox proportional hazard models were used to assess whether distributions of RFS, DFS, or OS differed with respect to the following clinical factors: age, extent of surgery, ER status, number of positive nodes, tumor size, Nottingham grade, HER2 status, prior exposure to estrogens ● Cross-validated HR in Cox proportional hazard models were obtained by the Faraggi-Simon method ● Multivariate Cox proportional hazard models and the likelihood-ratio test were applied to assess the contribution of the dichotomized ratio to the model ● In multivariate analysis tumor size, nodal status, tumor grade were included in model | <p>A high 2-gene expression ratio is associated with increased relapse and death in patients with node-negative ER positive breast cancer treated with tamoxifen.</p> |

Evidence Table 13. Clinical utility and validity, H/I ratio (continued)

| Study, year | Measure | Conclusions |
|------------------------|---|-------------|
| Goetz, 2006 (cont') | <p>Results:</p> <ul style="list-style-type: none"> • Nodal status, tumor size and Nottingham grade significantly associated with endpoints • All patient in the study: <ul style="list-style-type: none"> RFS, Univariate Cox 1.90 (CI 95% 1.25, 2.90) RFS, Univariate F-S 1.62 (CI 95% 1.06, 2.48) RFS, Multivariate Cox 1.65 (CI 95% 1.06, 2.57) RFS, Multivariate F-S 1.45 (CI 95% 0.93, 2.27) DFS, Univariate Cox 2.01 (1.36, 2.96) DFS, Univariate F-S 1.69 (1.14, 2.51) DFS, Multivariate Cox 1.77 (1.17, 2.66) DFS, Multivariate F-S 1.57 (1.04, 2.38) OS, Univariate Cox 1.95 (1.25, 3.07) OS, Univariate F-S 1.55 (0.98, 2.45) OS, Multivariate Cox 1.66 (1.04, 2.66) OS, Multivariate F-S 1.29 (0.81, 2.08) • Node negative patients in the study (n=130): <ul style="list-style-type: none"> RFS, Univariate Cox 2.22 (1.22, 4.05) RFS, Univariate F-S 1.99 (1.09, 3.63) RFS, Multivariate Cox 1.98 (1.07, 3.68) RFS, Multivariate F-S 1.73 (0.92, 3.25) DFS, Univariate Cox 2.41 (1.39, 4.18) DFS, Univariate F-S 2.12 (1.22, 3.68) DFS, Multivariate Cox 2.03 (1.15, 3.59) DFS, Multivariate F-S 1.77 (0.99, 3.16) OS, Univariate Cox 2.90 (1.47, 5.72) OS, Univariate F-S 2.35 (1.21, 4.58) OS, Multivariate Cox 2.40 (1.19, 4.84) OS, Multivariate F-S 2.01 (1.02, 3.99) | |

Evidence Table 13. Clinical utility and validity, H/I ratio (continued)

| Study, year | Measure | Conclusions |
|----------------------------|---|--|
| Jansen, 2007 ⁷² | <p>Context: Validation of the H/I signature by measuring expression levels normalized to a different set of control genes respect to Ma et al 2006.⁶¹</p> <p>Aims: to evaluate if the HOXB13-to-IL17BR expression ratio predicts response to tamoxifen, and/or cancer intrinsic aggressiveness</p> <p>End points: disease-free survival (DFS), progression free survival (PFS), post-relapse survival (PRS), and overall survival (OS);</p> <ul style="list-style-type: none"> ● Subsets of patients used from the total population (n=1252) for specific end points: <ul style="list-style-type: none"> DFS ER+, node negative primary breast cancer, no adjuvant therapy (N = 468); PFS in ER+ primary breast cancer, whose recurrence was treated with first-line tamoxifen (N = 193) <p>Methods:</p> <ul style="list-style-type: none"> ● Two-gene ratio as dichotomized or continuous variable ● Non-parametric methods: <ul style="list-style-type: none"> Spearman rank correlations for ER status, PR status, age; Kruskal - Wallis exact test for grade, size, LN status; Mann-Whitney <i>U</i> test for menopausal status; ● Univariate and multivariate Cox regression analysis to compute the hazard ratio ● In multivariate analysis the model for OS and DFS included age, menopausal status, tumor size, lymph node status, grade, and log ER and log PgR mRNA levels ● The model for PRS and PFS included age, menopausal status, DFS, site of relapse, and log ER and log PgR mRNA levels <p>Results:</p> <ul style="list-style-type: none"> ● All 1252 patients, associations with clinical factors, HOXB13: <ul style="list-style-type: none"> Grade, P < 0.001 ER status, P < 0.001 PR status, P < 0.001 ● All 1252 patients, associations with clinical factors, I17RB (both 3' and 5'): <ul style="list-style-type: none"> Age, P < 0.001; Menopausal status, P < 0.005 and P < 0.026 Tumor size P < 0.035 and P < 0.098 LN status, P < 0.001 Grade, P < 0.001 ER status, P < 0.001 PR status, P < 0.001 | <p>High HOXB13-to-IL17BR ratio expression levels associate with both tumor aggressiveness and tamoxifen therapy failure.</p> <p>The ratio was significantly associated with DFS and PFS in the specific subsets of patients.</p> <p>In multivariate analysis, the ratio was associated with a shorter DFS for node-negative patients only.</p> |

Evidence Table 13. Clinical utility and validity, H/I ratio (continued)

| Study, year | Measure | Conclusions |
|------------------------------|---|---|
| Jansen, 2007 (cont') | <ul style="list-style-type: none"> • Multivariate analysis, ER+, node negative, no adjuvant therapy (N = 468): Continuous ratio with 3' I17RB, DFS HR = 1.04; 95% CI = 1.01 to 1.08; P = 0.015 Continuous ratio with 5' I17RB, DFS HR = 1.05; 95% CI = 1.02 to 1.08; P = 0.004 Dichotomized ratio with 3' I17RB, DFS HR = 1.74; 95% CI = 1.17 to 2.59; P = 0.006 Dichotomized ratio with 5' I17RB, DFS HR = 1.61; 95% CI = 1.08 to 2.41; P = 0.019 • Multivariate analysis, relapsing ER+, tamoxifen treated (N = 193): Continuous ratio with 3' I17RB, PFS HR = 1.07; 95% CI = 1.04 to 1.11; P < 0.001 Continuous ratio with 5' I17RB, PFS HR = 1.05; 95% CI = 1.02 to 1.08; P = 0.004 Optimal dichotomized ratio with 3' I17RB, PFS HR = 2.97; 95% CI = 1.82 to 4.86; P < 0.001; Standard dichotomized ratio with 3' I17RB, PFS HR = 1.95; 95% CI = 1.39 to 2.73; P < 0.001; Optimal dichotomized ratio with 5' I17RB, DFS HR = 3.31; 95% CI = 2.05 to 5.34; P < 0.001; Standard dichotomized ratio with 5' I17RB, DFS HR = 2.12; 95% CI = 1.52 to 2.97; P < 0.001; | |
| Jerevall, 2007 ⁶³ | <p>Context: Validation of the H/I ratio signature. Expression levels were normalized to b-actin.</p> <p>Aims: to evaluate if the HOXB13-to-IL17BR expression ratio can predict the benefit of 5 years versus 2 years of tamoxifen treatment of postmenopausal patients;</p> <p>End points: Recurrence-free survival (RFS), defined as the time from diagnosis to local, regional, or distant recurrence or death due to breast cancer; OS, defined as the time elapsed from diagnosis to the date of death due to breast cancer;</p> <p>Methods:</p> <ul style="list-style-type: none"> • The relationships between grouped variables were analyzed with the chi-squared test, or the chi-squared test for trend when required; • Spearman's rank order correlation was used when comparing HOXB13 and IL17BR levels; • Survival curves were calculated using the Kaplan-Meier method and analyzed with the log rank test. • Multivariate analysis of recurrence and mortality rates was performed with Cox proportional hazard model; • A Cox model was also used to test for the interaction between the two-gene ratio and the benefit from prolonged duration of tamoxifen treatment; • The following variables were included: HOXB13:IL17BR, treatment (5 vs. 2 years), lymph node status (N+ vs. N-), tumor size (> 20 mm vs. > 20 mm), PgR status (PgR+ vs. PgR), and an interaction variable (HOXB13:IL17BR x treatment); | <p>The ratio or HOXB13 alone can predict the benefit of endocrine therapy, with a high ratio or a high expression rendering patients less likely to respond.</p> <p>Lower expression of IL17BR, but not HOXB13, was correlated to several factors related to poor prognosis, IL17BR might be an independent prognostic factor in breast cancer.</p> |

Evidence Table 13. Clinical utility and validity, H/I ratio (continued)

| Study, year | Measure | Conclusions |
|------------------------|--|-------------|
| Jerevall, 2007 (cont') | <p>Results:</p> <ul style="list-style-type: none"> ● The ratio is significantly correlated to: <ul style="list-style-type: none"> tumor size, P = 0.003 ER, P < 0.001 PR, P < 0.001 HER2, P = 0.003 NHG, P < 0.001 Ploidy, P < 0.001 S-phase, P = 0.005 ER, HER2, S-phase and NHG correlations are mostly due to IL17BT PR and ploidy correlation have contribution from both genes ● Post-menopausal ER+ patients, low ratio was associated with a benefit from a prolonged tamoxifen treatment (P = 0.021; in KM analysis for RFS), mainly due to the low expression of HOXB13 genes (P = 0.010, in KM analysis for RFS) ● Postmenopausal ER+ patients (n=179), multivariate Cox proportional hazard model analysis: <ul style="list-style-type: none"> Recurrence Rate (5y vs 2y), low ratio: 0.39 (CI 95% = 0.17–0.91), P value = 0.030 Test for interaction: P value = 0.035 Recurrence Rate (5y vs 2y), HOXB13: 0.37 (CI 95% = 0.17–0.83), P value = 0.015 Test for interaction: P value = 0.018 ● Postmenopausal ER+, node negative, patients (n=134), multivariate Cox proportional hazard model analysis: <ul style="list-style-type: none"> Recurrence Rate (5y vs 2y), low ratio: 0.27 (CI 95% = 0.10–0.72), P value = 0.0087 Test for interaction: P value = 0.014 Recurrence Rate (5y vs 2y), HOXB13: 0.30 (CI 95% = 0.12–0.73), P value = 0.0083 Test for interaction: P value = 0.0081 | |

Evidence Table 13. Clinical utility and validity, H/I ratio (continued)

| Study, year | Measure | Conclusions |
|------------------------|---|--|
| Ma, 2004 ⁶⁴ | <p>Context: This study used the two-gene ratio signature for the development and preliminary validation of the of the two-gene ratio signature prognostic value in patients with early stage breast cancer treated with tamoxifen. The training set was analyzed by microarray (n=60) and by RT-PCR (n=59); the validation set (n=20) was evaluated by RT-PCR (3' probes used)</p> <p>End points: DFS was calculated from the date of diagnosis.</p> <p>Methods:</p> <ul style="list-style-type: none"> ● Two-sample t test; ● Receiver Operating Characteristic (ROC) analysis; Statistical test of significance of ROC curves by the DeLong method ● Univariate and multivariate logistic regression; All predictors (tumor size, PGR, ERBB2, 2 gene-ratio) used as continuous variables ● Kaplan-Meier survival analysis and log-rank test <p>Results:</p> <ul style="list-style-type: none"> ● Microarray data, training set, two-genes ratio, Univariate logistic regression: OR = 10.17 (95% CI 2.9–35.6, p Value = 0.0003) ● Microarray data, training set, two-genes ratio, Multivariate logistic regression: OR = 7.3 (95% CI 2.1–26.3, p Value = 0.002 This outperforms the other variables ● Classification results in the validation set (RT-PCR data): 16/20 correctly classified Correct prediction probability = 0.01 (95% CI = 56%–94%), by exact binomial test; 9/10 disease free patients correctly classified 7/10 relapse patients correctly classified ● Kaplan-Meier survival analysis, log-rank test, RT-PCR on the training set: P value = 0.000058 ● Kaplan-Meier survival analysis, log-rank test, RT-PCR on the validation set: P value = 0.002 | <p>2-gene ratio predicts tumor recurrence in the setting of tamoxifen therapy.</p> |

Evidence Table 13. Clinical utility and validity, H/I ratio (continued)

| Study, year | Measure | Conclusions |
|------------------------|--|---|
| Ma, 2006 ⁶¹ | <p>Context: Development of the HOXB13:IL17BR two-gene index, through validation of the HOXB13:IL17BR two-gene ratio signature.</p> <p>Aim: to evaluate if the two-gene expression index (HOXB13:IL17BR) could predict outcome in patients treated and untreated with tamoxifen monotherapy</p> <p>End points: Relapse-free survival (RFS), defined as the time from initial diagnosis to any recurrence. Optimal threshold for dichotomization was identified and applied in the analysis</p> <p>Methods:</p> <ul style="list-style-type: none"> ● Spearman rank correlation was used to assess association between the HOXB13:IL17BR index and the other prognostic factors ● Cox proportional hazards regression and Kaplan-Meier analyses were used to examine the associations between gene expression indices and RFS ● Using untreated patients, cut-point selection was tuned to obtained the smallest log-rank P value using values between the 10th and 90th percentile; Cut-point selection was obtained from a training set and validated on a test set; Such cut-off was then used in treated patients <p>Results:</p> <ul style="list-style-type: none"> ● Interaction of HOXB13:IL17BR index with other clinical factors in ER+ patients: <ul style="list-style-type: none"> Node status, $r = 0.13$, P value = 0.0015 HER2, $r = 0.17$, P value < 0.0001 S-phase, $r = 0.15$, P value = 0.0004 ER, $r = -0.23$, P value < 0.0001 PR, $r = -0.26$, P value < 0.0001 ● Univariate Cox regression analysis: <ul style="list-style-type: none"> All patients (n=852), Interquartile HOXB13:IL17BR HR = 1.47 $p = 9e-06$ Node positive (n = 239), Interquartile HOXB13:IL17BR HR = 1.19 $p = 0.1$ Node negative (n = 613), Interquartile HOXB13:IL17BR HR = 1.64, $p = 2e-05$ ● Kaplan-Meier survival analysis results, ER+, node negative patients: <ul style="list-style-type: none"> Untreated patients, Test set (n=103), P value 0.0012 Treated patients, (n=122), P value =0.026 ● Multivariate Cox Regression Analysis; ER+ node negative, untreated test set and tamoxifen treated patients (n = 225), dichotomized HOXB13:IL17BR index (high versus low): <ul style="list-style-type: none"> HR = 3.9 (95% CI = 1.5 to 10.3) p value = 0.007 ● Two-gene index on a continuous scale, 5-year recurrence risk for untreated patients: <ul style="list-style-type: none"> Index of -2.0 = 15% (95% CI, 9.8% to 20.5%) Index of +2.0 = 36% (95% CI, 26.5% to 45.2%) | <p>The HOXB13:IL17BR index was only significant in node-negative patients.</p> <p>Higher HOXB13:IL17BR index was associated with a higher risk of relapse.</p> <p>Two-gene index was a significant predictor of clinical outcome in ER+, node-negative, patients irrespective of tamoxifen therapy.</p> |

Evidence Table 13. Clinical utility and validity, H/I ratio (continued)

| Study, year | Measure | Conclusions |
|--------------------------|---|---|
| Reid, 2005 ⁶⁹ | <p>Context: Validation of the H/I ratio signature using an independent cohort of 58 ER+ patients treated with tamoxifen. Re-analysis of the original microarray data and of an independent data set</p> <p>End points: DFS</p> <p>Methods:</p> <ul style="list-style-type: none"> ● RT-PCR measurement of HOXB13 and IL17BR gene expression (5' probes used) ● Univariate logistic regression ● Area Under the receiver-operating-Characteristic curve (AUC) ● Two-sample <i>t</i> test and Mann-Whitney test <p>Results:</p> <ul style="list-style-type: none"> ● Univariate logistic regression: odds ratio: <ul style="list-style-type: none"> HOXB13 OR = 1.04, 95% CI = 0.92 to 1.16, P = 0.54 IL17BR OR = 0.69, 95% CI = 0.40 to 1.20, P = 0.18 HOXB13/IL17BR OR = 1.30, 95% CI = 0.88 to 1.93, P = 0.18 ● Similar results by the other methods ● About Ma 2004 validation data: <ul style="list-style-type: none"> 9/10 disease free patients were correctly classified; Correct prediction probability 0.02 (95% CI = 52%–99%) 7/10 relapse patients correctly classified; Correct prediction probability = 0.34 (95% CI = 35%–93%) | <p>Although the proposed predictive model is very appealing the use of the two-gene ratio signature in an independent population yielded statistically non-significant results.</p> <p>The authors failed to confirm the association of the ratio with response to tamoxifen on their cohort (which is however different in terms of clinical characteristics from the original Ma, 2004⁶⁴ cohort). The authors also failed to classify patients using Discriminant Linear Analysis on two published data sets, including the Ma, 2004⁶⁴ original series.</p> |

ER = estrogen receptor; OS = overall survival; DFS= disease free survival; HR= hazard ratio; CI= confidence interval; RFS = recurrence free survival; RT-PCR= reverse transcriptase polymerase chain reaction; RNA= ribonucleic acid; HER-2 = human epidermal growth factor receptor 2; PFS = progression LN = lymph node; mRNA= messenger ribonucleic acid; NHG= Nottingham histologic grade; ROC= receiver operating characteristic; PGR= progesterone receptor gene; AUC= area under the receiver operating characteristic curve